UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE MATEMÁTICA

INSTITUTO SUPERIOR DE CIÊNCIAS DO TRABALHO E DA EMPRESA
DEPARTAMENTO DE FINANÇAS

Stochastic Modelling of Non-Stationary Financial Assets

Joana Ribeiro Estevens

Mestrado em Matemática Financeira

Dissertação orientada por:
Professor Doutor João Pedro Silva Brito Boto
Doutor Pedro Gonçalves Lind

2016

Acknowledgements

I would like to thank for the amazing opportunity of finishing my master's dissertation in Germany under the programme Erasmus+ sponsored by the European Union. This research was partially supported by the grant they gave me.

I would also like to express my sincere gratitude to my colleagues from the department of Statistical-Physics from the University of Osnabrück who gave me a warm welcome in Germany and provided me with useful comments and remarks about my work. A special thanks to Professor Philipp Maass not only for his insights but also for the hard questions that allowed be to expand my work.

This dissertation could not have been completed without the guidance, support and encouragement that I received from my advisors, Professor João Boto and Doctor Pedro Lind. A very special and heartfelt thank you for both of them who always kept the office door open and were always available whenever I needed it.

I thank my fellow colleague Paulo Rocha for the useful discussions, the data sharing and for his infinite patient in helping me overcome my difficulties during this period.

Finally, I would like to thank my family and friends who were always there when I needed. To my parents that supported me in each and every imaginable way throughout my entire life. To Cristiana who has always given me the moral support that I needed. To my partner in life, my everyday support, Gonçalo, I thank you for giving me more than I could have asked for.

Abstract

In this dissertation, we propose a framework to model the evolution of the non-stationary time series of the volume-price from 2000 companies from the New York Stock Exchange (NYSE). It was shown that for each 10 minutes window, the distribution that best fits our data is the log-normal distribution. Since the volume-price series is non-stationary, we treat the parameters ϕ and θ of the log-normal distribution as stochastic variables. We assume that all the time dependency of the volume-price series is included in the parameters, $\phi(t)$ and $\theta(t)$. Therefore, by describing the evolution of ϕ and θ we are able to model the volume-price series.

Our analysis decomposes the parameters evolution into average daily patterns, $\overline{\phi}$ and $\overline{\theta}$, and fluctuations around these patterns, ϕ' and θ'.

The daily patterns $\overline{\phi}$ and $\overline{\theta}$ are easily modelled using cubic curves. The fluctuations ϕ' and θ' are modelled using Langevin equations. We show that both parameters fluctuations, ϕ' and θ', are weakly correlated, which leads us to consider two different models. One, where we assume that the variables are not correlated and therefore yields two independent Langevin equations for our data. And another, two-dimensioned, where we take into account the correlation between the two variables ϕ' and θ' which leads to the derivation of a system of coupled Langevin equations. We used the first approach because most of the times, the simplest models are the ones used in the real world applications. However, we know that the variables are indeed correlated. Thus, we also followed the other approach where we took into account this correlation since we expect it to yield better results.

Finally, we use the system of Langevin equations and the cubic curves describing the daily patterns to reconstruct two synthetic time series statistically reproducing the evolution of the parameters ϕ and θ, and show that they unequivocally determine the expected value of each statistical moment of the log-normal distribution of the volume-price. By comparing the probability density function of these expected values using the ϕ and θ time series obtained from our model versus the empirical ones, we arrive to the conclusion that our model describes well the first moments of the volume-price time series. This framework proposed by us is general enough to be applied to other fields of study where non-stationary stochastic variables need to be modelled, like biology, medicine, geology, among others.

The data from the NYSE was collected directly from Yahoo Finance between January 2011 and April 2014, with a sampling frequency of 10 minutes, which give us a total of 976 days.

Keywords: Stochastic Differential Equations, Non-Stationarity, Time Series, Brownian Motion.

Resumo

Nesta dissertação vamos propor um modelo que nos permite estudar a série temporal não estacionária do volume-preço de 2000 empresas cotadas no New York Stock Exchange. Estes dados foram recolhidos diretamente do Yahoo Finance entre 27/01/2011 e 06/04/2014, com uma frequência de amostragem de 10 minutos. O volume-preço num dado instante corresponde ao produto entre o preço de determinada ação pelo volume transacionado da mesma. Esta é uma variável muito importante na matemática financeira pois incorpora a interação existente entre o volume e o preço de uma ação. Por exemplo, volumes altos têm tendência a originar preços altos, enquanto volumes baixos estão habitualmente associados a preços mais baixos. Para além disso, quando uma pessoa decide comprar/vender uma ação, o preço da mesma não é a única variável importante. Também é necessário ter em conta o volume, visto que este está associado à liquidez do título, isto é, ao quão fácil ou difícil é comprar/vender a ação. Para além disso, a distribuição do volume-preço dá-nos informação sobre a quantidade de capital que está a ser transacionada no mercado.

Apesar da série temporal do volume-preço ser não estacionária, esta segue uma forma funcional constante ao longo do tempo. Já foi mostrado em trabalhos anteriores [1] que, para cada janela temporal de 10 minutos, o modelo que melhor explica o volume-preço das 2000 empresas da bolsa de Nova Iorque é uma log-normal com parâmetros ϕ e θ. A média do logaritmo do volume-preço é representada pelo parâmetro ϕ enquanto o desvio-padrão do logaritmo do volume-preço é representado pelo parâmetro θ.

É sabido da estatística clássica que é possível chegar a uma fórmula fechada para o n-ésimo momento de uma log-normal e que, sabendo a expressão de todos os momentos de uma distribuição, podemos chegar à sua função densidade de probabilidade usando transformadas de Fourier. Portanto, o nosso objetivo de estudar a série não estacionária do volume-preço resume-se a estudar as séries estacionárias dos parâmetros ϕ e θ.

A nossa análise decompõe a evolução dos parâmetros ϕ e θ na soma dos seus padrões médios diários, $\overline{\phi}$ e $\overline{\theta}$, com as flutuações em torno destes parâmetros, ϕ' e θ'. Vamos modelar estas duas partes separadamente e utilizando abordagens diferentes.

Modelar os padrões diários $\overline{\phi}$ e $\overline{\theta}$ é um procedimento simples, visto que ambos se podem descrever por uma função cúbica. A modelação das flutuações já tem que ser mais cuidadosa pois estamos perante séries temporais com um comportamento fortemente estocástico. Como tal, vamos modelar estas séries através de equações diferenciais estocásticas, também conhecidas como equações de Langevin. É importante referir que nós dividimos as séries temporais ϕ e θ em duas partes porque não é possível modelar dados que contenham algum tipo de periodicidade através de equações de Langevin. Portanto, tivemos que retirar a parte periódica dos nossos dados. Utilizámos a transformada rápida de Fourier para nos certificarmos de que não persistia nenhum tipo de periodicidade nas nossas séries temporais das flutuações.

Começámos por modelar as flutuações assumindo que as séries ϕ' e θ' são independentes uma da outra. Apesar de isto não se verificar na realidade, o coeficiente de correlação entre as séries é muito baixo pelo que podemos supor, para um modelo mais simples, que é aproximadamente zero. Para além disso, na maior parte das situações da vida real, os modelos mais simples são preferíveis face aos mais complexos, visto que é mais fácil implementá-los e interpretar os seus resultados. Como tal, nós extraímos dos nossos dados os coeficientes que regem as duas equações de Langevin que modelam as flutuações dos nossos dados. Em ambos os casos, o coeficiente da parte determinística da equação, $D^{(1)}$, corresponde a uma função linear da variável em estudo, enquanto o coeficiente da parte estocástica, $D^{(2)}$, é uma função quadrática.

Após esta primeira abordagem com um modelo mais simples, passámos para um modelo que incorpora a interação entre ϕ' e θ'. Vamos descrever estas flutuações através de um sistema de duas equações Langevin acopladas. Agora o coeficiente da parte determinística, $D^{(1)}$, é na verdade um vetor de duas funções enquanto o coeficiente da parte estocástica, $D^{(2)}$, é uma matriz simétrica com três funções distintas.

Visto que temos um sistema de equações que nos permite descrever a evolução de ϕ' e θ', podemos construir as nossas próprias séries temporais das flutuações. Somando estas séries teóricas ao padrão médio obtido pelas funções cúbicas, ficamos com séries temporais teóricas para o ϕ e para o θ . Por outras palavras, estes são os valores de ϕ e θ que o nosso modelo prevê. Comparámos estes valores teóricos das séries do ϕ e do θ com os valores empíricos que já tínhamos. Chegámos à conclusão de que o nosso modelo é capaz de explicar muito bem a evolução da série temporal do ϕ, visto que a densidade da série teórica e da série empírica são quase coincidentes. Contudo, a série do θ já não é tão bem explicada pelo nosso modelo. Isto já era esperado visto que, como o ϕ é um momento de primeira ordem, então é mais facilmente modelado do que um momento de segunda ordem. Notar que o θ é a raiz quadrada do segundo momento centrado.

Como já vimos anteriormente, é possível chegar à fórmula fechada para o o n-ésimo momento da distribuição log-normal que depende apenas de ϕ e θ. Após termos esta fórmula, podemos substituir os valores de ϕ e θ pelos que obtivemos através do nosso modelo e obtemos a série temporal do n-ésimo momento teórico. Nós calculámos as séries dos primeiros quatro momentos teóricos. Para percebermos se os nossos resultados estavam de acordo com a realidade, fizemos um processo semelhante para as séries do ϕ e do θ empíricas. Finalmente, comparámos as funções densidade de probabilidade de ambas as séries e percebemos que o nosso modelo descreve muito bem os primeiros momentos da série do volume-preço. Mais uma vez, isto também já era esperado visto que os momentos de ordens superiores estão mais dependentes do θ do que aqueles de ordens inferiores.

Há inúmeros modelos na literatura que nos permitem estudar séries temporais. Uma pergunta pertinente seria o porquê de termos escolhido esta análise de Langevin em detrimento dos outros modelos. Uma particularidade muito interessante sobre o nosso modelo é que ele não nos permite apenas descrever a evolução temporal da série do volume-preço. Para além disso, também podemos chegar a uma equação às derivadas parciais, de tipo Fokker-Plank, que nos dá a evolução da função densidade de probabilidade do volume-preço. A dedução desta equação não foi efetuada no âmbito desta tese, mas é um excelente ponto para ser desenvolvido em trabalhos futuros.

Para terminar, o modelo que nós propusemos nesta tese pode ser aplicado à bolsa de valores de Nova Iorque para calcular medidas de risco como o *Value at Risk*. Uma das questões que se levantaram a partir da realização deste trabalho é se este modelo é adequado para fazer previsões sobre a evolução do volume-preço. Este seria um bom ponto de partida para trabalhos futuros. Para além disso, este modelo é suficientemente geral para poder ser aplicado a séries temporais de outras áreas científicas, como por exemplo no estudo da variabilidade cardíaca na fisiologia ou no estudo de séries sísmicas na geologia. Este trabalho proporcionou-nos uma visão bastante aprofundada do estudo das séries temporárias não estacionárias e permitiu-nos propor uma metodologia que será bastante útil em várias áreas.

Palavras-Chave: Equações Diferenciais Estocásticas, Não Estacionariedade, Séries Temporais, Movimento Browniano.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this dissertation we are going to explore a way of understanding the evolution of the volume-price in the stock market. We use data from 2000 companies in the New York Stock Exchange (NYSE). We are going to do this using stochastic differential equations.

Volume-price is an important variable in mathematical finance since it incorporates the interaction between volume and price. For instance, the volume has as an important role in the assets' price. High volumes trigger high prices and low volumes are associated with low prices. Moreover, while we need to know the price of an asset if we want to sell or buy it at the right time, we should also know the volume that is being transitioned in the market, since high volumes reflect a good market liquidity, i.e. it is easy to sell or buy the asset. Finally, the distribution of the volume-price provides information about the capital that is being transitioned in the market.

We want to study the temporal evolution of the volume-price which is a non-stationary random variable. If it would be a stationary variable, it would be very easy to solve this problem. In the literature we have numerous ways to model stationary variables [2]. However, since we have a non-stationary random variable, it seems almost impossible to fit a good model to it. Recently, it was shown [1] that although the distribution of the volume-price is non-stationary, the series has a constant functional shape (log-normal) throughout time. The parameters of this log-normal distribution are time dependent and they are stationary. We will study these stationary parameters in order to describe the evolution of the non-stationary time series of the volume-price.

Figure 1.1 shows a simple scheme of the idea that will be explored in this dissertation. The starting points of our study are the volume and price time series of 2000 companies in the New York Stock Exchange (NYSE). This data was collected directly from Yahoo Finance, with a sampling frequency of 10 minutes, starting in January 27th 2011 and ending in April 6th 2014, which yields a total of 976 days ($\approx 10^5$ data points). The data preprocessing was done in a previous work and can be found in Refs. [3, 4]. We can see this time series for just one company in Figure 1.1a. Multiplying the two series yields the volume-price time series. For the same time interval, we have a sample of approximately 2000 companies. Thus, for each 10-minutes snapshot, we have 2000 observations of volume-price, one for each company.

It was shown [1] that the log-normal distribution had the best fit to this data. We can see a 10-minutes snapshot in Figure 1.1b: the dots represent the empirical probability density function (PDF) of the logarithm of the volume-price time series for the 2000 companies in that particular 10 minutes. The solid line is the PDF of a normal distribution with mean and standard deviation

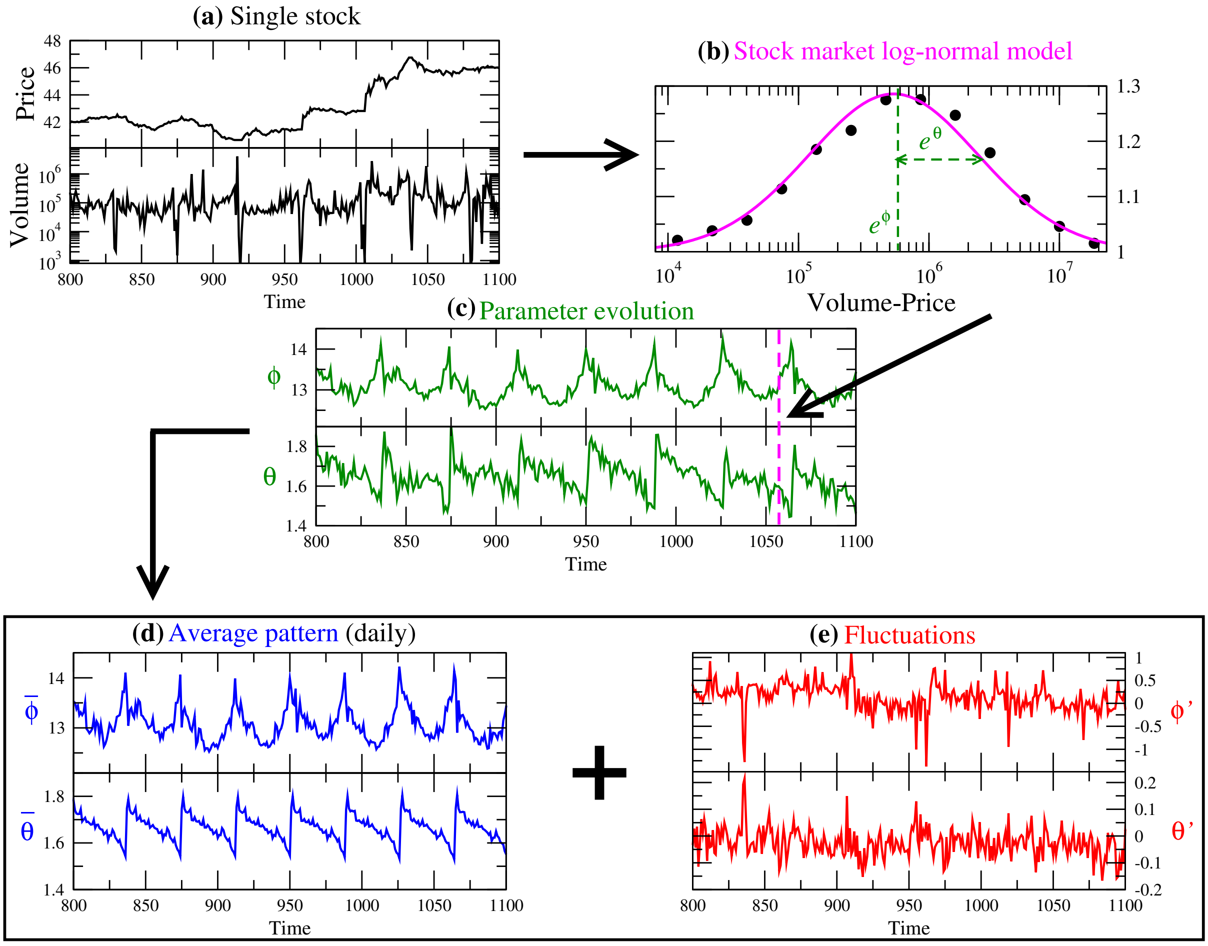

Figure 1.1: We start with the price and volume series of 2000 different companies. In (a) we can see the volume and price series for just one of the companies. Multiplying the two series yields the volume-price series which follows a log-normal distribution with parameters ϕ and θ. In (b) we can see the empirical density represented by the dots and the adjusted log-normal to the data, solid line, for a particular window of 10 minutes. Each 10-minutes window yields a log-normal with different parameters. Thus, we will have (c) a time series for the parameter ϕ and another one for θ. Each time series can be decomposed in (d) a daily pattern and (e) fluctuations around this pattern. We will describe the evolution observed in (c) by analysing this components separately.

equal to the ones of the volume-price logarithm. However, the series is non-stationary: another 10-minutes window yields a log-normal with different parameters. This means that our volume-price series follows a constant functional form throughout time (log-normal) but the parameters of this distribution are themselves stochastic variables.

We are going to study the evolution of these parameters illustrated in Figure 1.1c. For each 10-minute window, we will have a value to the mean ϕ and to the standard deviation θ of the volume-price logarithm. We can also see in this figure that both time series appear to have a daily pattern. Because of this, we are going to split our analysis in two parts: one considering the average daily patterns, Figure 1.1d, and another one with the fluctuations around this pattern, Figure 1.1e. The original series is the sum of these two parts.

We will propose a framework to describe the evolution observed in Figure 1.1c, modelling average behaviour and stochastic contributions separately. We are going to extract stochastic differential equations for the parameters ϕ and θ following a recent framework [5]. By knowing how the fluctuations from the ϕ and θ time series behave, we will be able to describe the evolution of the non-stationary series of volume-price. The main goal of this dissertation is to model the original non-stationary time series of volume-price, s.

In this introductory chapter we stated the problem that is going to be addressed by this dissertation, the importance of this subject and what we did, in general lines, to solve this problem as we can see in Figure 1.1. In Chapter 2, we will do a quick review of the main concepts which are necessary to understand the subsequent chapters. In Chapter 3, we will explain the preprocessing of the ϕ and θ time series. After that, we explore various aspects of these series and we check if it is possible the analysis of these time series. In Chapter 4 we will analyse both time series independently and coupled. In Chapter 5, we will use the results presented in the previous chapters to derive the equations that govern the volume-price non-stationary time series. Finally, the discussions and conclusions are presented in Chapter 6.

Chapter 2

State of the art

2.1 Stationary Stochastic Processes

A stochastic process (X_t) is a family of random variables indexed by t belonging to some index set T. The index set T can be any abstract set. However, in order to make it simple, we take it as time, and hence T can be taken as $\mathbb{R}$ or some subset of $\mathbb{R}$. In this case, we often call the stochastic process (X_t) a time series. When we are trying to explain real world phenomena with stochastic processes, we can notice that we only have a incomplete sample path of the process. However, we want to figure out the probability structure from this data, i.e. the finite dimensional distributions. If we do not make more assumptions, this is basically an impossible task. In this section, we are going to represent by $\mathcal{B}(\mathbb{R}^n)$ the σ-algebra generated by the family of open intervals from $\mathbb{R}^n$.

Definition 1 (Finite Dimensional Distributions). *Let $(X_t)_{t\geq 0}$ be a stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$. The probabilities*

$$P_{X_{t_1},X_{t_2},...,X_{t_n}}(B) = P((X_{t_1}, ..., X_{t_n}) \in B) \tag{2.1}$$

where $n \in \mathbb{N}, t_1, ..., t_n \in \mathbb{R}, 0 \leq t_1 < t_2 < ... < t_n < \infty, B \in \mathcal{B}(\mathbb{R}^n)$, i.e. B is a Borel measurable set of $\mathbb{R}^n$, are called finite dimensional distributions.

Finite dimensional distributions completely characterize the probability structure of a stochastic process since they involve the knowledge of all marginal distributions. However, if you do not make further simplifications or study special cases we will not be able to get tractable models for stochastic processes. One special case that is easier to study is a stationary process. A stationary stochastic process in its simplest form assumes that it is in equilibrium. In fact the probability structure of the process does not change in time.

Definition 2 (Strict Stationarity). *A stochastic process $(X_t)_{t\geq 0}$ is strictly stationary if for any $t_1, t_2, ..., t_n$ and h, the joint distributions of $(X_{t_1}, ..., X_{t_n})$ and $(X_{t_1+h}, ..., X_{t_n+h})$ are identical, that is*

$$P_{X_{t_1},X_{t_2},...,X_{t_n}}(B) = P_{X_{t_1+h},X_{t_2+h},...,X_{t_n+h}}(B) \tag{2.2}$$

where $n \in \mathbb{N}, t_1, ..., t_n \in \mathbb{R}, 0 \leq t_1 < t_2 < ... < t_n < \infty, B \in \mathcal{B}(\mathbb{R}^n)$

It is important to notice that a strictly stationary process has a probability structure which is invariant under a time shift. This means the random variables X_t and X_{t+h} have the same distribution. Therefore, if the moment exists then $E(X_t^n) = E(X_{t+h}^n)$. In particular, a strictly stationary process will have constant mean and variance. Besides that, the covariance function of a strictly stationary time series depends only on the time interval between the time points, not on the specific location of the points along the time axis.

Strict stationarity is a very restrictive property defined on all of the finite dimensional distributions which is very difficult to verify empirically. There are weaker forms of stationarity, expressed in terms of the moments which are often sufficient to construct very useful processes.

Definition 3 (mth-order Stationarity). *The process X_t is said to be stationary up to order m, if for any $t_1, t_2, ..., t_n$ and for all $m_1, m_2, ..., m_n$ such that $\sum_{i=1}^{n} m_i \leq m$, all product moments exist $E(X_{t_1}^{m_1}...X_{t_n}^{m_n})$ and for any $h \geq 0$,*

$$E(X_{t_1}^{m_1}...X_{t_n}^{m_n}) = E(X_{t_1+h}^{m_1}...X_{t_n+h}^{m_n}) \quad . \tag{2.3}$$

Second-order stationarity, which is also often called weak stationarity or covariance stationarity, is fundamental in studying large classes of processes. Strictly stationary processes do not need to be m^{th}-order stationary. However, when moments up to order m exist, then a strictly stationary process will also be a m^{th}-order stationary. Since the knowledge of full infinite sequence of moments under certain conditions define the finite dimensional distributions, loosely speaking, strict stationarity corresponds to m^{th}-order stationarity in the limit as $m \to \infty$.

2.2 Brownian Motion and White Noise

Until the nineteenth century, it was commonly thought that if we could collect all the initial data then we would predict the future with certainty. This is known as the Laplace's dream: we would be able to create a model of the universe which would be completely deterministic [6]. We now know that this is not true. The limited predictability may arise in the form of fluctuations due to interactions with the environment.

In 1828, the botanist Robert Brown observed and tried to describe the irregular movement of pollen, suspended in water [7]. This movement is now known as the Brownian movement and is attributed to the collisions of the pollen with the water molecules, resulting in a diffusion of the pollen in the water. His work was largely ignored by the scientific community at his time.

In 1900, independently from Brown's work, L. Bachelier [8] derived the law governing the position w_t at time t of a single grain of pollen performing a one-dimensional Brownian motion starting at $a \in \mathbb{R}$ at time t = 0:

$$P_a\{W_t \in dx\} = p(t, a, x)dx \quad , \quad with \quad p(t, a, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-a)^2}{2t}} \tag{2.4}$$

where $p(t, a, x)$ is the solution of the heat equation:

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial a^2}, \tag{2.5}$$

In 1905 Albert Einstein addressed the Brown's question without having knowledge of Bachelier work [9]. He predicted what the Brownian motion should be. These predictions are nowadays a part of one of the formal definitions of Brownian motion [10].

However, it was only in 1923 that Nobert Wiener proved the existence of Brownian motion and set down a firm mathematical foundation for its analysis. In his honour, the mathematical formulation of the Brownian motion is know as the Wiener process. Before we give the mathematical definition of a Brownian motion, it is important to understand the concept of a filtration.

Definition 4 (Filtration). *Let $X : [0, +\infty[\times \Omega \to \mathbb{R}$ be a stochastic process on a probability space Ω with a σ-algebra $\mathcal{F}$. For each $0 \leq t < \infty$, we define a σ-algebra $\mathcal{F}_t$ as the sigma algebra generated by the variables $X_s,\ 0 \leq s \leq t$:*

$$\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t) \tag{2.6}$$

If $0 \leq s < t$, then $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$. Such a family of σ-fields $\mathcal{F}_t : 0 \leq t < \infty$ is called a filtration of $\mathcal{F}$.

This filtration is generated by the process X_t since it contains all of its past at each time t. In an intuitive way, it is the smallest filtration available to study the process X_t: is contains all the information regarding the past of the process, but only that information. Now we can see what a Brownian motion is:

Definition 5 (Brownian Motion). *A Brownian motion is a real valued, continuous stochastic process $(W_t)_{t\geq 0}$, with the following properties:*

- $W_0 = 0$ *almost surely.*
- *independent increments: If $s \leq t$, $W_t - W_s$ is independent of $\mathcal{F}_s = \sigma(W_u : 0 \leq u \leq s)$.*
- *stationary increments: If $s \leq t$, $W_t - W_s$ and $W_{t-s} - W_0$ have the same probability law since they follow a normal distribution with mean zero and variance $t - s$.*

Using the Kolmogorov theorem [10] it is possible to prove that the function $t \mapsto W(t)$ is almost surely continuous. However, the randomness allows Brownian motion to also be nowhere differentiable as we are going to see. However, we clarify the concept of limit of a stochastic variables that we are going to use here.

Definition 6 (Almost Sure Convergence). *We say that X_n converges to X with probability one (almost surely) if $P\left(\lim_{n\to+\infty} X_n = X\right) = 1$. That is, for almost every $\omega \in \Omega$ except for a set of $\omega \in \Omega$ with measure 0, the sequence $X_n(\omega)$ converges to $X(\omega)$.*

Theorem 1 (Non-differentiability of the Brownian motion). *Almost surely, Brownian* *motion is nowhere differentiable. Furthermore, almost surely, for all t, either*

$$\lim_{\sup_h \to 0} sup \frac{W_{t+h} - W_t}{h} = +\infty \tag{2.7}$$

or

$$\lim_{\inf_h \to 0} sup \frac{W_{t+h} - W_t}{h} = -\infty \quad . \tag{2.8}$$

Since the sample paths of a Brownian motion are nowhere differentiable with probability 1, a stochastic process of the form $(\xi_t)_{t\geq 0}$ with:

$$\xi_t = \frac{dW_t}{dt} = W_t' \;\; or \;\; dW_t = \xi_t dt \tag{2.9}$$

cannot be introduced by taking the almost sure limit in a difference quotient. Nevertheless, it is possible to arrive to a definition via an integral [11]. In order to approach this definition, let $g(t)$ be any function with a continuous derivative $g'(t)$ in the interval $[a, b]$ and $t_0, t_1, ..., t_n$ a sequence of numbers satisfying:

$$a = t_0 < t_1 < ... < t_n = b \;\; and \;\; \Delta t_i = t_{i+1} - t_i; i = 0, 1, 2..., n-1$$

Then, the stochastic integral $\int_a^b g(t)dW_t$ is defined as the almost sure limit:

$$\int_a^b g(t)dW_t = \lim_{n\to+\infty,\ max\Delta t_i \to 0} \sum_{i=0}^{n-1} g(t_i)(W_{t_i+\Delta t_i} - W_{t_i}) = \tag{2.10}$$

$$= \lim_{n\to+\infty,\ max\Delta t_i \to 0} g(b)W_b - g(a)W_a - \sum_{i=0}^{n-1} W_{t_{i+1}} \frac{g(t_i + \Delta t_i) - g(t_i)}{\Delta t_i} \Delta t_i \tag{2.11}$$

The stochastic integral is an almost sure limit of a sum of normally distributed variables. Therefore, it also has a normal distribution. It is important to notice that we are going to give a more general definition of the stochastic integral later, after we have introduced more concepts. Now we just want to motivate the following definition by taking the almost sure limit on both sides of the previous equation.

Definition 7 (White Noise). *Let $(W_t)_{t\geq 0}$ be a Brownian motion. A stochastic process $(\xi_t)_{t\geq 0}$ is called a white noise if it satisfies, for any function g(t), with a continuous derivative $g'(t)$ in $[a, b], a < b$, the following relationship:*

$$\int_a^b g(t)\xi_t dt = g(b)W_b - g(a)W_a - \int_a^b W_t g'(t)dt \quad . \tag{2.12}$$

It is important to notice that in the left side of Equation (2.12) we are writing an abusive notation (although common and inoffensive) since ξ_t does not exist as a function whereas it is defined as a linear functional in the space of continuous differentiable functions through the right side of Equation (2.12).

If W_t had a first derivative in order to time, then $\xi_t = \frac{dW_t}{dt}$ would satisfy the relationship in the previous definition. Therefore, the white noise ξ_t can be interpreted as a generalized derivative of the Brownian motion W_t.

If $s \leq t$ we know that $W_t - W_s \sim N(0, t-s)$. Therefore $E(W_t) = E(W_t - W_0) = 0$ and $var(W_t) = var(W_t - W_0) = t - 0 = t$. If $s < t$, $E(W_s W_t) = E(W_s(W_s + W_t - W_s)) = E(W_s^2) + E(W_s)E(W_t - W_s) = var(W_s) + 0 = s$. We can replicate these calculations if $t < s$ and we would get $E(W_s W_t) = t$. Therefore, $E(W_s W_t) = min\{s, t\}$.

Since $E(W_t) = 0$ and $E(W_s W_t) = min\{s, t\}$, we know that $\mathrm{Cov}(W_s, W_t) = min\{s, t\}$. Using this result for the Brownian motion and realizing that $\xi_t = \frac{dW_t}{dt}$, it is easy to see that $\mathrm{Cov}(\xi_s, \xi_t) = 0$. Therefore, for $s \neq t$ there is no correlation between ξ_s and ξ_t, even if we make the difference $|s - t|$ very small.

As a result, white noise can be seen as the most random stochastic process and this is why it is so often used to model random noise in various fields like electronics, engineering, econometrics, finance, among others. But also because of this property, the white noise cannot exist in the real world.

It is not plausible for any physical phenomena to generate uncorrelated ξ_s and ξ_t even when $|s - t|$ is very small. This mismatch from physical reality appears in its mathematical representation as well: one cannot define a continuous white noise process with non-zero, finite variance. The variance of a white noise process is undefined [12].

In practice, a weakly stationary stochastic process $(\xi_t)_{t \geq 0}$ can approximately be considered white noise if the covariance between ξ_t and $\xi_{t+\tau}$ tends extremely fast to 0 with increasing τ.

The denomination white noise comes from the spectral theory of stationary random processes that states the white noise has a power spectrum which is uniformly distributed over all frequencies (like white light).

The range of application of Brownian motion goes far beyond a study of microscopic particles in suspension and included modelling of stock prices, thermal noise in electrical circuits, limiting behaviour in queuing and random perturbations in a variety of other physical, biological, economic and management systems. Moreover, integration with respect to Brownian motion gives us a unifying representation for a large class of diffusion processes.

Diffusion processes represented this way exhibit a rich connection with the theory of partial differential equations. In particular, to each diffusion process there corresponds a second order parabolic equation which governs the transition probabilities of the process.

Bachelier not only was the pioneer in modelling the Brownian motion using stochastic processes, but he was also the first one to build a theory for the fluctuations of the stock market using advanced mathematics. Besides that, he uncovered the Markovian property of the Brownian motion. However, his work went largely unknown for nearly a century.

2.3 Markov Process

The Brownian motion is an example of a Markov process. This type of processes are very important since we can derive results assuming the Markovian property which we would not be able to get without it. The Langevin Analysis that we are testing in this thesis is one example of this. We can only use this approach if our time series are Markovian [13] . Before we define a Markov process, it is important to understand the concept of conditional expectation. It is known from basic probability theory that the conditional probability of an event $A \in \mathcal{F}$ given $B \in \mathcal{F}$ and $P(B) > 0$ is the real number:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.13}$$

Using this simple concept, we can create a new probability measure:

Theorem 2 (Conditional probability of an Event). *Let us consider a probability space $(\Omega, \mathcal{F}, P)$ and $B \in \mathcal{F}$ with $P(B) > 0$. Thus, the application $Q_B : \mathcal{F} \to [0, 1]$ defined by $Q_B(A) = P(A|B)$ is a new probability measure in $(\Omega, \mathcal{F})$ and it is called the conditional probability of event A given event B.*

And now we are able to define the conditional expectation of a stochastic random variable, generalizing Equation (2.13):

Definition 8 (Conditional Expectation of a Stochastic Variable given an event). *Let $B \in \mathcal{F}$ with $P(B) > 0$ and let $\xi : \Omega \to \mathbb{R}$ be a stochastic integrable variable, i.e., $\mathbb{E}(|\xi|) < \infty$. The conditional expectation of ξ given an event B, $\mathbb{E}(\xi|B)$, is the real number:*

$$\mathbb{E}(\xi|B) = \frac{\mathbb{E}(\mathbb{1}_B \xi)}{P(B)} \tag{2.14}$$

where

$$\mathbb{1}_B(x) = \begin{cases} 1, & \text{if } x \in B \\ 0, & \text{if } x \notin B \end{cases} \tag{2.15}$$

Definition 9 (Conditional Expectation of a Stochastic Variable given a σ-algebra). *Let $\mathcal{G}$ be a sub- σ-algebra from $\mathcal{F}$ and let $\xi : \Omega \to \mathbb{R}$ be a stochastic integrable variable. The conditional expectation of ξ given the σ-algebra $\mathcal{G}$ is a stochastic integrable variable, $\mathbb{E}(\xi|\mathcal{G})$ that satisfies the following properties:*

- *$\mathbb{E}(\xi|\mathcal{G})$ is measurable with respect to $\mathcal{G}$.*
- *If $A \in \mathcal{G}$ then $\mathbb{E}(\mathbb{1}_A \mathbb{E}(\xi|\mathcal{G})) = \mathbb{E}(\mathbb{1}_A \xi)$*

Since we are now familiar with the concept of conditional expectation, we can easily understand that $\mathbb{E}(X|\mathcal{F}_t)$ is the best estimate of X based on observations of the process up to time t. The properties of conditional expectations with respect to filtrations define various types of stochastic processes. One of the most important ones is the Markov process.

Definition 10 (Markov Process). *Let us consider a stochastic process $(X_t)_{t \geq 0}$ in a probability space $(\Omega, \mathcal{F}, P)$. We say that $(X_t)_{t \geq 0}$ is a Markov process if it satisfies the following property (Markov property): to all the functions $f : \mathbb{R} \to \mathbb{R}$, Borel measurable and bounded and $0 \leq s \leq t < \infty$:*

$$\mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s) \tag{2.16}$$

If X is any random variable, then $E(X \mid \mathcal{F}_t)$ is the best estimate of X based on observations of the process up to time t. Intuitively, one interprets Equation (2.16) as follows: given the *present* of the process, the *future* is independent of its *past*. A Markov process only cares about its present state, and has no memory of how it got there.

Markov processes are very useful in applied mathematics for several reasons. Four of the more important reason are:

- Many real world phenomena can be modeled by a Markov process.
- Usually the input needed for the application of a Markov process is more easily given than for other non-Markovian processes.
- There are various computer algorithms for numerical simulations for Markov processes.
- All stochastic processes that have independent increments also have the Markov property.

Theorem 3 (Brownian Motion as a Markov Process). *Let us consider a probability space $(\Omega, \mathcal{F}, P)$. A Brownian motion $(W_t)_{t\geq 0}$ is a Markov process.*

It is also possible to characterize a Markov process using its finite dimensional distributions. Let us consider the times:

$$0 \leq t_1 < t_2 < ... < t_m < t_{m+1} < ... < t_n$$

The conditional probability that $X_{t_i} = x_i$ for $m+1 \leq i \leq n$ given that $X_{t_i} = x_i$, for $1 \leq i \leq m$, is given by:

$$p(x_n, t_n; ...; x_{m+1}, t_{m+1} \mid x_m, t_m; ...; x_1, t_1) = \frac{p(x_n, t_n; ...; x_1, t_1)}{p(x_m, t_m; ...; x_1, t_1)} \tag{2.17}$$

We have a Markov process if these conditional densities depend only on the most recent time, which means:

$$p(x_{n+1}, t_{n+1} \mid x_n, t_n; ...; x_2, t_2; x_1, t_1) = p(x_{n+1}, t_{n+1} \mid x_n, t_n) \tag{2.18}$$

Therefore, we also have for a Markov process:

$$p(x_n, t_n; ...; x_2, t_2 \mid x_1, t_1) = p(x_n, t_n \mid x_{n-1}, t_{n-1})...p(x_2, t_2 \mid x_1, t_1) \tag{2.19}$$

It is now possible to deduce all joint finite dimensional probability densities of a Markov process X_t using only the transition density $p(x, t \mid y, s)$ and the probability density of its initial value, $p_0(y)$:

$$P(x,t) = \int_{-\infty}^{+\infty} p(x, t \mid y, 0) p_0(y) dy \tag{2.20}$$

The transition probabilities of a Markov satisfy the Chapman-Kolmogorov equation:

$$p(x, t \mid y, s) = \int_{-\infty}^{+\infty} p(x, t \mid z, r) p(z, r \mid y, s) dz \quad , \quad s < r < t \tag{2.21}$$

Intuitively this means that the probability of a transition from y at time s to x at time t is equal to the probability of the transition to z at an intermediate time r, multiplied by the probability of the transition from z at the time r to x at the time t, summed over all possible intermediate values z.

Definition 11 (Homogeneous Markov Process). *A Markov process $(X_t)_{t\geq 0}$ is (time) homogeneous if*

$$p(x,t \mid y,s) = p(x,t-s \mid y,0) \tag{2.22}$$

In other words, a Markov process is said to be time homogeneous if its transition probability is stationary. A homogeneous Markov process is essentially a Markov process with invariable stochastic properties under a time shift. The probability of a transition from y to x only depends on the time difference $t-s$. Therefore, we can write $p(x,t|y,s) = p(x,t-s|y)$. We can also rewrite the Chapman-Kolmogorov equation:

$$p(x,t \mid y) = \int_{-\infty}^{+\infty} p(x,t-s \mid z)p(z,s \mid y)dz \quad , \quad 0 < s < t \tag{2.23}$$

Theorem 4 (Brownian Motion as a Homogeneous Markov Process). *The Brownian motion $(W_t)_{t\geq 0}$ is a homogeneous Markov process with state space $\mathbb{R}$ and transition probability function given by*

$$p(x,t \mid y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} \qquad t > 0 \quad . \tag{2.24}$$

2.4 Stochastic Differential Equations

In Equation (2.9) we established that the white noise can be seen as the time derivative of the Brownian motion. In fact, the Brownian motion is the foundation for the constitution of an extensive class of Markov processes with continuous sample paths, called diffusion processes. The white noise was an example. But we can also have more sophisticated examples. For instance, we can add a mean drift to the white noise. This diffusion process can be described by the stochastic differential equation:

$$\frac{dX_t}{dt} = b(X_t) + \xi_t \tag{2.25}$$

where $b : \mathbb{R} \to \mathbb{R}$ is a given smooth function and ξ_t is a white noise. We can think about this equation as a white noise perturbed by a drift term b or as a deterministic ordinary differential equation perturbed by an additive white noise [13].

The white noise is the time-integral of the Brownian motion. Therefore, any differential equation with a white noise can be rewritten as an integral equation with a Brownian motion. We rewrite Equation (2.25) as:

$$X_t = X_0 + \int_0^t b(X(s))ds + W(t) \tag{2.26a}$$

or in the differential form

$$dX_t = b(X_t)dt + dW_t \tag{2.26b}$$

These are stochastic differential equations (SDE) which contain a white noise with a constant strength. In order to study SDE where the strength of the white noise depends on the solution, we will need to give a meaning to the expression:

$$\int_0^T f_t(\omega)dW_t(\omega) \tag{2.27}$$

where $(\Omega, \mathcal{F}, P)$ is a probability space and $(W_t)_{t\geq 0}$ is a Brownian motion.

One of the most important properties of the Brownian motion is that its paths are almost surely not differentiable at any point. Therefore, we cannot define the integral in Equation (2.27) as a Lebesgue-Stieltjes integral. Nevertheless, we are able to define this type of integral with respect to a Brownian motion and we will call them *stochastic integrals.*

To start, we will construct the integral for a set of processes called simple processes. After that, the definition of the integral will be extended to a more general class of processes by taking a limit [10].

Definition 12 (Simple Process). *$(S_t(\omega))_{0\leq t\leq T}$ is called a simple process if it can be written as*

$$S_t(\omega) = \sum_{i=1}^{p} \phi_i(\omega)\mathbb{1}_{]t_{i-1},t_i]} \quad . \tag{2.28}$$

where $0 = t_0 < t_1 < ... < t_p = T$ and ϕ_i is $\mathcal{F}_{t_{i-1}}$-measurable and bounded.

Definition 13 (Stochastic Integral of a Simple Process). *The stochastic integral of a simple process S defined as in Equation 2.28 is the continuous process:*

$$\int_0^T S_t\, dW_t := \sum_{i=1}^{n} \phi_i(W_{t_i} - W_{t_{i-1}}) \tag{2.29}$$

Now we want to extend the previous definition to a broader class of stochastic processes. We will define the stochastic integral for all the stochastic processes that belong to $\mathcal{H}^2([0,T])$, defined as:

Definition 14 (Stochastic Processes in $\mathcal{H}^2([0,T])$). *Given a time interval $[0,T]$, let $\mathcal{H}^2([0,T])$ be the class of stochastic processes $H = (H_t)_{t\in[0,T]}$ that satisfy the following conditions:*

- *$(H_t)_{t\in[0,T]}$ is a measurable process, i.e. the function $[0,T]\times\Omega \to \mathbb{R} : (t,\omega) \mapsto H_t(\omega)$ is measurable.*
- *$(H_t)_{t\in[0,T]}$ is adapted to the filtration .$(\mathcal{F}_t)_{t\geq 0}$, i.e. for all $t \in [0,T]$ the variable H_t is $\mathcal{F}_t$-measurable*
- *$E(\int_0^T H_t^2) < \infty$.*

For every process in $\mathcal{H}^2([0,T])$ there is a sequence of simple processes that converge to that one process. This result is going to let us define the stochastic integral for all the processes in $\mathcal{H}^2([0,T])$.

Theorem 5. *To every stochastic process $H \in \mathcal{H}^2([0,T])$, there is a sequence $(H^n)_{n\in\mathbb{N}}$ of simple stochastic processes that:*

$$\lim_{n\to\infty} E\left(\int_0^T (H_t - H_t^n)^2\right) = 0 \quad . \tag{2.30}$$

Therefore, it is easy to see that if we have a process $H \in \mathcal{H}^2([0,T])$ and a sequence $(H^n)_{n\in(\mathbb{N})}$ in the conditions of the previous theorem then $\int_0^T (H_t^n)_{n\in\mathbb{N}}\, dW_t$ also converges and the limit does not depend on the sequence $(H^n)_{n\in(\mathbb{N})}$. Hence, we are now ready to give a definition of the stochastic integral in $\mathcal{H}^2([0,T])$.

Definition 15 (Stochastic Integral). *For a given process $H \in \mathcal{H}^2([0,T])$,we will define the stochastic integral of Itô of H in $[0,T]$ as the almost sure limit of the sequence of stochastic integrals $\int_0^T (H_t^n)_{n\in\mathbb{N}}\, dW_t$, where $(H^n)_{n\in\mathbb{N}}$ is in the conditions of Theorem 5, i.e.*

$$\int_0^T H_t\, dW_t = \lim_{n\to\infty} \int_0^T H_t^n\, dW_t \tag{2.31}$$

Since we already know what a stochastic integral is, we can then generalize Equations (2.26) and define a SDE in a more general way. In Equations (2.26) we examined SDE that contains a white noise with a constant strength. Now we are going to define SDE where the strength of the white noise depends on the solution and we are going to allow the coefficients do depend explicitly on time.

Definition 16 (Stochastic Differential Equation). *Let us consider a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ and let $(W_t)_{t\geq 0}$ be a Brownian motion. A stochastic differential equation is an equation with the following form:*

$$X_t = \alpha + \int_{t_0}^t f(s, X_s)ds + \int_{t_0}^t g(s, X_s)dW_s \quad , \quad t \geq t_0 \geq 0 \tag{2.32}$$

where $f : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}, (t,x) \mapsto f(t,x)$, $g : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}, (t,x) \mapsto g(t,x)$ are measurable functions, α is a random variable and $(X_t)_{t\geq 0}$ is a stochastic process. The functions f and g are called the coefficients of the equation and α is the initial condition.

We can also write a SDE in the differential form:

$$\begin{cases} dX_t = f(t, X_t)dt + g(t, X_t)dW_t \quad , \quad t \geq t_0 \\ X_{t_0} = \alpha \end{cases} \tag{2.33}$$

A non rigorous explanation of Equation 2.33 (but a very useful one) is that for a little time increase $dt > 0$, the stochastic process $(X_t)_{t\geq 0}$ changes its value in a random quantity that is normally distributed with mean $f(t, X_t)dt$ and variance $g(t, X_t)g^T(t, X_t)dt$ and it is independent from the past of the process. This happens because the increments of the Brownian motion are independent and normally distributed with mean 0 and variance equal to the time increase. In this sense, the term $g(t, X_t)dW_t$ is used to model the perturbation of the random noise that affects the deterministic system $dX_t = f(t, X_t)dt$. In a physicist tradition we sometimes refer to a SDE as a Langevin equation. Both designations mean the same.

If we want to work with SDE, we not only have to know what the stochastic integral means, but we also have to know how to work with it. The traditional rules of calculus will not apply to the stochastic integral.

One of the key results that allows us to work with the stochastic integral is a new version of the chain rule called Itô's formula.

Theorem 6 (Itô's Formula). *Let $(X_t)_{t\geq 0}$ be a solution of Equation 2.33 and $h : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}, (t, x) \mapsto h(t, x)$ a function with continuous derivative with respect to t and continuous derivatives with respect to x until the second order. In another words, this means h is a function in $\mathcal{C}^2([0, +\infty[\times\mathbb{R})$.*

Then $(Y_t)_{t\geq 0}$ where $Y_t = h(t, X_t)$ is a stochastic process that satisfies the equation

$$dY_t = \left(\frac{\partial h}{\partial t}(t, X_t) + f(t, X_t)\frac{\partial h}{\partial x}(t, X_t) + \frac{1}{2}g^2(t, X_t)\frac{\partial^2 h}{\partial x^2}(t, X_t)\right) dt + g(t, X_t)\frac{\partial h}{\partial x}(t, X_t)dW_t \quad . \tag{2.34}$$

Now that we have already understood how the Itô's formula work in one dimension, we will generalize it to several dimensions [14].

Theorem 7 (Itô's Formula in d-dimension). *Let $\mathbb{X}_t = (X_t^1, ..., X_t^d)^T$ be a vector process satisfying:*

$$d\mathbb{X}_t = \mathbb{A}dt + \mathbb{H}d\mathbb{W}_t \tag{2.35}$$

where $\mathbb{W}_t$ is a multi-dimensional independent Brownian motion defined as $\mathbb{W}_t = (W_t^1, ..., W_t^n)^T$, $\mathbb{A} = (a_1(t, \mathbb{X}_t), ..., a_d(t, \mathbb{X}_t))^T$ is the drift vector and $\mathbb{H}$ is the diffusion matrix:

$$\mathbb{H} = \begin{bmatrix} h_{11}(t, \mathbb{X}_t) & ... & h_{1n}(t, \mathbb{X}_t) \\ \vdots & \ddots & \vdots \\ h_{d1}(t, \mathbb{X}_t) & ... & h_{dn}(t, \mathbb{X}_t) \end{bmatrix} \tag{2.36}$$

Let $h : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ be a given bounded function in $\mathcal{C}^2([0, +\infty[\times\mathbb{R}^d)$. Let $Y_t = h(t, \mathbb{X}_t)$. Then,

$$dY_t = \left(\frac{\partial h}{\partial t} + \sum_{i=1}^{d} a_i \frac{\partial h}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d}\sum_{p=1}^{n} h_{jp}h_{ip}\frac{\partial^2 h}{\partial x_i \partial x_j}\right) dt + \sum_{j=1}^{d}\sum_{p=1}^{n} h_{jp}\frac{\partial h}{\partial x_j}dW_t^p \tag{2.37}$$

2.5 The Fokker-Planck Equation

We have already seen that the Brownian motion is a time-homogeneous Markov process with a transition density that satisfies Equation (2.5). Fokker and Planck derived this differential equation for the distribution function describing the Brownian motion and it is now known as the Fokker-Planck equation for the Brownian motion. However, we can generalize this to other stochastic processes. In fact, the Fokker-Planck equation is an equation of motion for the distribution function of fluctuating macroscopic variables, i.e. it describes the evolution through time of the probability density function of macroscopic variables. By solving the Fokker-Planck equation one obtains distribution functions from which any averages of macroscopic variables are obtained by integration.

In order for us to present the Fokker-Planck equation, it is necessary to be familiar with the concept of a diffusion process which is a special case of a Markov process with continuous sample functions which serve as probability-theoretic models of physical diffusion phenomena [15]. The simplest and oldest example of a diffusion process is the Brownian motion. There are different approaches to the class of diffusion processes. We are going to define them in terms of the conditions on the transition probabilities $p(X_t \in B | X_s = x)$.

Definition 17 (**Diffusion Process**). *A diffusion process is a Markov process* $X_t,\ t_0 \leq t \leq T$, *with values in* $\mathbb{R}^d$ *and almost sure continuous sample functions whose transition probability* $p(X_t \in B|X_s = x)$ *satisfies the following three conditions for every* $s \in [t_0, T[,\ x \in \mathbb{R}^d$ *and* $\epsilon > 0$*:*

- $$\lim_{s \to t} \frac{1}{t-s} \int_{|y-x|>s} p(X_t = y|X_s = x)dy = 0 \ . \tag{2.38}$$

- *There exists a* $\mathbb{R}^d$*-valued function* $D^{(1)}(s, x)$ *such that:*

$$\lim_{s \to t} \frac{1}{t-s} \int_{|y-x| \leq s} (y-x) p(X_t = y|X_s = x)dy = D^{(1)}(s, x) \ . \tag{2.39}$$

- *There exists a* $d \times d$ *matrix-valued* $D^{(2)}(s, x)$ *such that:*

$$\lim_{s \to t} \frac{1}{t-s} \int_{|y-x| \leq s} (y-x)(y-x)' p(X_t = y|X_s = x)dy = D^{(2)}(s, x) \ . \tag{2.40}$$

The functions $D^{(1)}$ and $D^{(2)}$ are called the coefficients of the diffusion process. In particular, $D^{(1)}$ is called the drift vector and $D^{(2)}$ is called the diffusion matrix. $D^{(2)}$ is non-negative-defined. Now we can proceed to see what a Fokker-Planck Equation is.

Theorem 8 (**Fokker-Planck Equation**). *Let* $X_t,\ t_0 \leq t \leq T$,*denote a d-dimensional diffusion process that satisfies the conditions in Definition 17 and which possesses a transition density* $p(X_t = y|X_s = x)$*. If the derivatives:*

$$\frac{\partial p}{\partial t},\ \frac{\partial}{\partial y_i}\left(D_i^{(1)}(t, y)p\right),\ \frac{\partial^2}{\partial y_i \partial y_j}\left(D_{ij}^{(2)}(t, y)p\right)$$

exist and are continuous functions, then for fixed s and x such that $s \leq t$, the transition probability $p(X_t = y|X_s = x)$ is a fundamental solution of the Fokker-Planck equation:

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial y_i} \left(D_i^{(1)}(t, y)p \right) - \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial y_i \partial y_j} \left(D_{ij}^{(2)}(t, y)p \right) = 0 \tag{2.41}$$

The Fokker-Planck equation can be used to get the probability density function associated with a SDE. In fact, the Itô's formula gives us a simple way of deriving the Fokker-Planck equation [13]. If we do this, we will see that the coefficients of the SDE are related to the coefficients of the Fokker-Planck equation.

Analytical solutions to the Fokker-Planck equations can only be reached in very special case. Most of the times we cannot find an analytical solution. However, it is possible to use other methods of solutions like simulation methods, transformation of the Fokker-Planck equation in a Schrödinger equation, numerical integration methods among others.

One of the main goals of this dissertation is to determine the coefficients $D^{(1)}$ and $D^{(2)}$ that govern the Fokker-Planck equation of the probability density function of the fluctuations of the parameters of the log-normal distribution of volume-price.

As an example of application of the Fokker-Planck equation, we will study the oldest example of a stochastic differential equation which describes the Brownian motion of a particle under the influence of friction but no other force field and it is known as the Ornstein-Uhlenbeck process:

Definition 18 (Ornstein-Uhlenbeck Process). *The Ornstein-Uhlenbeck process is the univariate continuous Markov process X_t that evolves with time t according to the following stochastic differential equation:*

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t \tag{2.42}$$

where $\theta, \mu, \sigma > 0$ and $(W_t)_{(t \geq 0)}$ is a Brownian motion.

We can solve Equation (2.42) and we get a solution that is the sum of a deterministic behaviour plus a stochastic term :

$$X_t = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \sigma \int_0^t e^{-\theta(t-s)})dW_s \tag{2.43}$$

The probability density function, $p(X_t = x|X_s = y)$, of the Ornstein-Uhlenbeck process satisfies the Fokker-Planck equation:

$$\frac{\partial p}{\partial t} - \frac{\partial}{\partial x} \left(\theta(x - \mu)p \right) - \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(\sigma^2 p \right) = 0 \tag{2.44}$$

In order to simplify the results, we are going to assume $\mu = 0$ and $D = \frac{1}{2}\sigma^2$. Then the solution for this Fokker-Planck equation is:

$$p(X_t = x|X_s = y) = \sqrt{\frac{\theta}{2\pi D(1 - e^{-2\theta(t-s)})}} exp \left(-\frac{\theta(x - e^{-\theta(t-s)})y^2}{2D(1 - e^{-2\theta(t-s)})} \right) \tag{2.45}$$

2.6 The Black-Scholes Model and its limitations

Although Bachelier had already applied results from stochastic calculus to the finance world, this methodology only started to be widely used when Fischer Black, Myron Scholes [16] and Robert Merton [17] addressed the problem of pricing and hedging an European option on a non-dividend paying stock in 1973. This is one of the most important problems in Financial Mathematics. Black and Scholes worked independently from Merton but arrived to the same conclusions. The Nobel Prize in Economics for 1997 was awarded to Merton and Scholes. If Black had been alive in that year, he would have shared the prize.

Although their model have some problems which have been pointed out throughout the years, it is widely employed as a useful approximation. However, a proper application requires a good understanding of its limitations. Blindly following the model exposes the user to unexpected risk.

The Black-Scholes model assumes that the asset price, S_t, follows a geometric Brownian motion:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{2.46}$$

where W_t is a standard Brownian motion, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. This stochastic differential equation tells us that the price variation of the asset in the time interval dt follows a normal distribution with mean μdt and variance $\sigma^2 dt$. Equation (2.46) is the main assumption of the Black-Scholes model. Besides that, we also assume short selling is allowed, there are no transactions costs, the assets are perfectly divisible and pay no dividend, there are no arbitrage opportunities, trading takes place continuously in time and the risk-free rate is constant and the same for all maturities. From all these assumptions, we can derive the time-t fair value of an European-style option (call or put) on an asset with spot price S_t [10].

Despite the importance of the Black-Scholes model, various authors have showed that its underlying assumptions do not agree with the market's reality. For example, the geometric Brownian motion cannot explain the negative skewness and the high kurtosis that are usually seen in the empirical asset return distributions [18]. Another example is the negative correlation between stock returns and realized volatility, known as *leverage effect* [19]. However the most noticed drawback in the Black-Scholes model is the inverse relationship between the implied volatility and the strike price, known as *volatility smile* [20].

Other models were developed in order to overcome these problems. The Constant Elasticity of Variance (CEV) model was developed by Cox in 1975 [21]. It is consistent with the *leverage effect* and with the *volatility smile*. The CEV model assumes that the asset price, S_t, follows the following diffusion process:

$$dS_t = \mu S_t dt + \sigma S_t^{\frac{\alpha}{2}} dW_t \tag{2.47}$$

where α is the constant of elasticity that controls the relationship between volatility and price. When $\alpha < 2$ we have the *leverage effect*, i.e. the observed variance of stock returns will be inversely related with the asset's price: it will increase as the asset's price decreases and decrease when the price increases.

Another important and useful alternative to the Black-Scholes model is the Heston model [22]. This is one of the most well known stochastic volatility models. It assumes that the asset price, S_t, is governed by the following stochastic differential equation:

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^{(1)} \tag{2.48}$$

where W_t is a standard Brownian motion and $\mu \in \mathbb{R}$. The instantaneous variance of the asset's returns is assumed to follow another stochastic differential equation:

$$dv_t = a(b - v_t)dt + \sigma\sqrt{v_t} dW_t^{(2)} \tag{2.49}$$

where b is the long term mean of v_t, $a \geq 0$ is the speed of mean reversion, i.e., the rate at which v_t reverts to b and σ is the volatility of the variance process. Moreover, the parameters have to obey the Feller condition so the process v_t is strictly positive:

$$\frac{2ab}{\sigma^2} > 1 \tag{2.50}$$

The Brownian motions in Equations (2.48) and (2.49) are correlated, with correlation ρ.

2.7 From Stochastic Volatility to Superstatistics

Superstatistics is a branch of statistics aimed to the study of non-linear and non-equilibrium systems. Complex systems often show a behaviour which can be regarded as a superposition of different dynamics[23]. Superstatistics uses the superposition of multiple statistical models to explain the complex system hence the prefix "super".

One example of a superstatistics is the Heston model. We saw that this model assumes that the asset price follows a stochastic differential equation (SDE), in particular Equation (2.48). However, the parameter v_t of SDE follows another SDE, Equation (2.49). Thus, we have a superposition of statistics that explain the evolution of the asset price S_t.

In our case, we have a non-stationary time series: the volume-price series. This is a complex system that we cannot describe using ordinary statistics. We will also use an ensemble of different statistics.

It was proved [1] that the volume-price series follows a log-normal distribution throughout time as we can see Equation (3.1). If the volume-price time series would be stationary we would not need superstatistics. However, the parameters ϕ and θ are not constants. They follow another statistical model. Thus, we are going to study the evolution of these parameters with the aim of uncovering the evolution of volume-price.

2.8 The Langevin Analysis

We are interested in modelling the parameters ϕ and θ of the log-normal distribution of volume-price. We are going to do this using stochastic differential equations (SDE) since we are going to find the functions $D^{(1)}$ and $D^{(2)}$ governing the Fokker-Planck equation for the probability density function of the parameters. The approach that we are going to follow was introduced in 1997 [24] and reviewed by Friedrich et al. in a paper from 2011 [5].

We want to describe the log-normal parameters with a stochastic differential equation of the type:

$$\frac{dX}{dt} = D^{(1)}(X,t) + g(X,t)\xi_t \quad , \tag{2.51}$$

where X is a vector which contains our parameters and ξ_t is a white noise. The Langevin analysis allow us to extract directly from the data the vector of functions $D^{(1)}(X)$ and the matrix of functions $D^{(2)}(X) = g(X)g^T(X)$.

A stochastic process described by Equation (2.51) can be modeled by stochastic evolution laws that relate the state vectors X(t) at times $t_i, t_{i+1} = t_{i+\tau}, t_{i+2} = t_{i+2\tau}, ...$, for small but finite values of τ. Here, we deal with the SDE that are defined by the following discrete time evolutions:

$$X(t_{i+1}) = X(t_i) + D^{(1)}(X(t_i), t_i)\tau + g(X(t_i), t_i)\sqrt{\tau}\xi(t_i) \quad . \tag{2.52}$$

This discrete SDE must be considered in the limit $\tau \to 0$. We are going to explain how the discrete time processes are related to Equation (2.51).

In order to motivate the discrete time approximations, we integrate the Equation (2.51) over a finite but small time increment τ.

$$X(t+\tau) = X(t) + \int_t^{t+\tau} D^{(1)}(X(s), s)ds + \int_t^{t+\tau} g(X(s), s)\xi(s)ds \tag{2.53}$$

$$\approx X(t) + D^{(1)}(X(t), t)\tau + \int_t^{t+\tau} g(X(s), s)\xi(s)ds \tag{2.54}$$

These are the quantities for which a statistical characterization can be given. We are going to interpret the integral for the wildly fluctuating stochastic quantities $\xi(t)$ in the Itô sense:

$$\int_t^{t+\tau} g(X(s), s)\xi(s)ds \approx g(X(t), t)\int_t^{t+\tau} \xi(s)ds = g(X(t), t)\sqrt{\tau}\eta(t) \tag{2.55}$$

where $\eta(t)$ is a stochastic variable with a standard Gaussian distribution since $W_t = \int_0^t \xi(s)ds$ and $W_{t+\tau} - W_t \sim \mathcal{N}(0, \sqrt{\tau})$.

We have just discussed processes described by stochastic equations. Now we are going to summarize the corresponding statistical description needed to the Langevin analysis.

$D^{(1)}$ and $D^{(2)}$ are the drift vector and the diffusion matrix in Equation (2.41). By considering the Itô's definitions of the stochastic integrals, the coefficients $D^{(1)}$ and $D^{(2)}$ of the Fokker–Planck equation and the coefficients of the Langevin equation are related. They are defined according to the following equations, where $\langle X \rangle$ is the mean of the variable X:

$$D_i^{(1)}(x,t) = \lim_{\tau \to 0} \frac{1}{\tau} < X_i(t+\tau) - x_i >|_{X(t)=x} \tag{2.56a}$$

$$D_{ij}^{(2)}(x,t) = \lim_{\tau \to 0} \frac{1}{\tau} < (X_i(t+\tau) - x_i)(X_j(t+\tau) - x_j) >|_{X(t)=x} \tag{2.56b}$$

These equations are the discrete versions of Equations (2.39) and (2.40). They show us that the drift vector and the diffusion matrix are determined as the first and second moments of the conditional probability distributions in the small time limit.

We will know describe the Langevin analysis which allows us to get the drift vector and the diffusion matrix directly from the data:

- The time series are represented in a state space, i.e., the set of values that a process can take.
- The state space is partitioned into a set of small bins.
- For each bin α, located at point x_α of the partition we consider the quantity:

$$x(t_j+\tau) = x(t_j) + D^{(1)}(x(t_j),t_j)\tau + g(x(t_j),t_j)\sqrt{\tau}\xi(t_j) \quad , \tag{2.57}$$

  where the points $x(t_j)$ are taken from the bin located at x_α.

The drift vector assigned to the bin located at x_α is determined as:

$$D^{(1)}(x_\alpha,t) = \lim_{\tau\to 0}\frac{1}{\tau}M^{(1)}(x_\alpha,t,\tau) \tag{2.58}$$

on the conditional moment:

$$M^{(1)}(x_\alpha,t,\tau) = \frac{1}{N_\alpha}\sum_{x(t_j)\in\alpha}[x(t_j+\tau)-x(t_j)] \tag{2.59}$$

where N_α is the number of points contained in the bin α.
The diffusion matrix is estimated by:

$$D^{(2)}(x_\alpha,t) = \lim_{\tau\to 0}\frac{1}{\tau}M^{(2)}(x_\alpha,t,\tau) \tag{2.60}$$

on the conditional moment:

$$M^{(2)}(x_\alpha,t,\tau) = \frac{1}{N_\alpha}\sum_{x(t_j)\in\alpha}([(x(t_j+\tau)-x(t_j)) - \tau D^{(1)}(x_j,t)]^2 \tag{2.61}$$

2.9 Statistical Tests

In this dissertation we are going to use some statistical tests that we present here in this section.

In order to apply the Langevin Analysis, it is necessary that the time series of the fluctuations, ϕ' and θ', are Markovian. Most of the physical phenomena can only be considered Markovian if we take a sufficient large time step. Einstein recognized this for the Brownian motion. He discussed the smallest time scale for which the Brownian motion can be seen as a Markovian stochastic process [25]. We call this time step the Markov length, τ_M.

We will use the Wilcoxon rank-sum test [26] to compare the conditional probabilities:

$$p(x_1, \tau_1 | x_2, \tau_2) \;\text{ and }\; p(x_1, \tau_1 | x_2, \tau_2; x_3, \tau_3) \; . \tag{2.62}$$

We will compute the value of t/t_0, where $t_0 = \sqrt{\frac{2}{\pi}}$ and $t = \frac{|Q - <Q>|}{\sigma(Q)}$ where Q denotes the total number of inversions in the Wilcoxon test. Q is a Gaussian distributed variable with mean value $< Q >$ and standard deviation $\sigma(Q)$. Therefore t is the absolute value of a Gaussian-distributed random variable with mean value zero and standard deviation one. The expected value of t, where averaging is done with respect to x_2 should be $\sqrt{\frac{2}{\pi}}$. Therefore, values of t/t_0 close to 1, indicates the data has the Markovian property.

Shapiro and Wilk's [27] W-test is a wide used and powerful test of departure from normality. It tests the null hypothesis that the sample $x_1, ..., x_n$ comes from a gaussian distributed population. The test statistics is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x}^2)} \tag{2.63}$$

where $x_{(i)}$ is the i-th order statistics, i.e., the i-th smallest number in the sample, $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and:

$$(a_1, ..., a_n) = \frac{m^T V^{-1}}{\sqrt{m^T V^{-1} V^{-1} m}} \tag{2.64}$$

where $m = (m_1, ..., m_n)^T$ and m_i, for $i = 1, ..., n$ is the expected value of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics.

Chapter 3

Getting to know the data

In previous works, it was shown that, to the volume-price at this time scale and with this data, the log-normal distribution had the best fit to the data among other four statistical models [28]. The probability density function of the log-normal is:

$$p(s) = \frac{1}{s\sqrt{2\pi}\theta} exp\left[-\frac{(\log s - \phi)^2}{2\theta^2}\right] \tag{3.1}$$

If volume-price is log-normal distributed with parameters ϕ and θ then the volume-price logarithm has a normal distribution with mean ϕ and standard deviation θ. This relationship is easily observed if we plot a log-normal variable in a logarithmic scale, like we did in Figure 1.1b. In this section, we are going to study the time series of the parameters ϕ and θ of this distribution.

3.1 Outliers and Daily Patterns

We start by removing the outliers from our data. An outlier is an observation that is distant from the other ones. The outliers may be the result of measuring errors. In order for us to have a robust model we should discard the outliers. After we have studied our data series, we decided to consider as outliers all the data points which do not lie within five standard deviations of the mean. By doing this, we removed 33 data points from our original series.

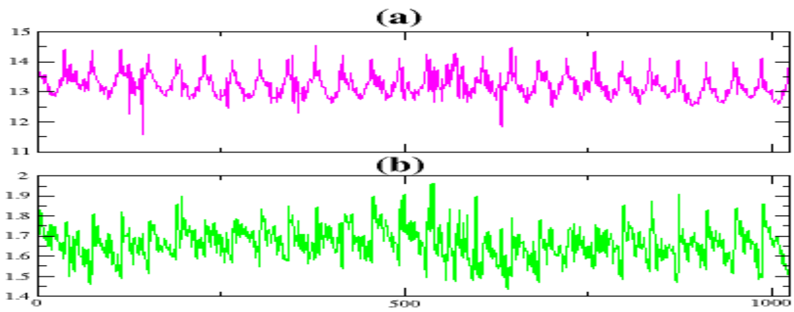

Figure 3.1: The first 27 trading days of the (a) ϕ and (b) θ time series after we have removed the outliers.

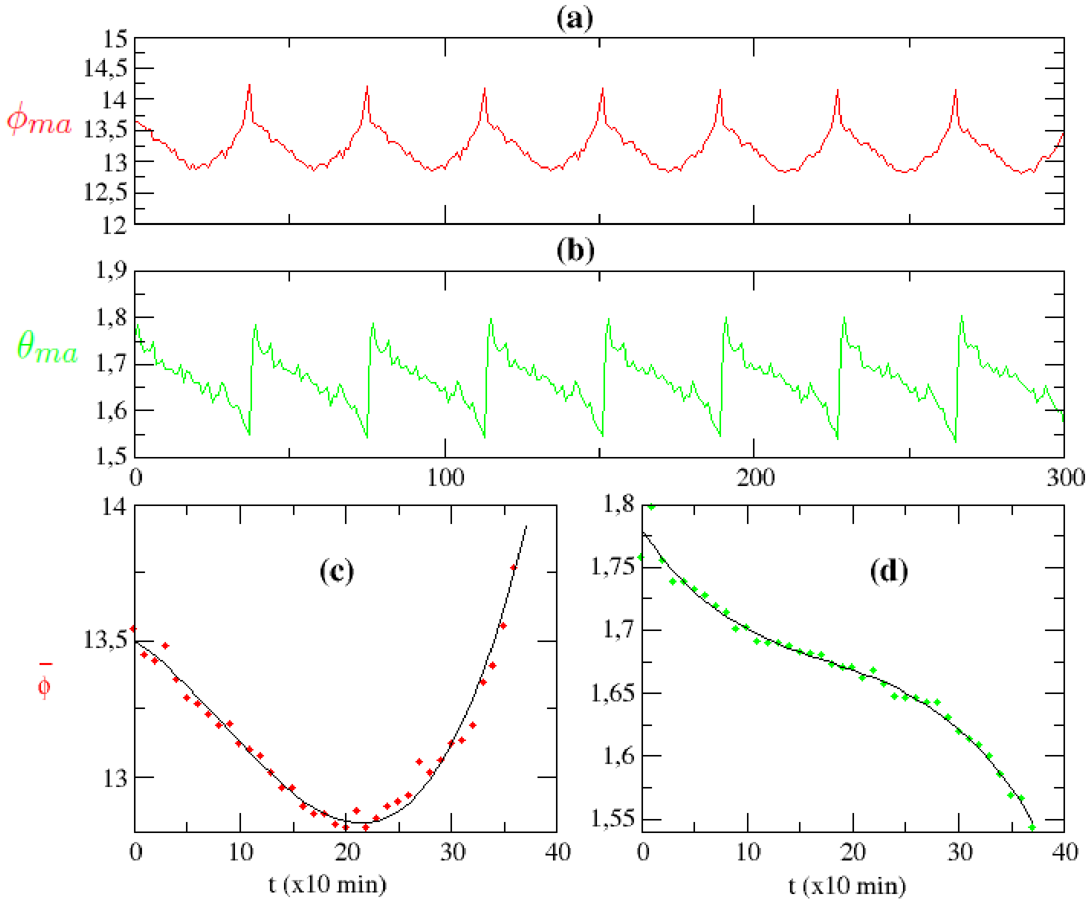

Figure 3.2: (a) ϕ_{ma} represents the 20-day moving average pattern of parameter ϕ over a period of 9 days and (b)θ_{ma} represents the same for θ. (c) The average over all trading days of parameters ϕ, i.e. $\overline{\phi}$, and (d) $\overline{\theta}$, with the respective fitting functions.

We can see the result from this process in Figure 3.1. We can also see that we clearly have a daily pattern in both time series. In order to apply the Langevin analysis described in Chapter 2, we will have to remove this pattern in order to obtain the time series of the fluctuations since the Langevin analysis does not support any kind of periodicity in the data. The original series ϕ and θ are the sum of the average daily pattern with the fluctuations around this pattern.

In Figure 3.2 we can see that both the average of ϕ and θ are described by the following cubic curves:

$$\overline{\phi}(t_d) = a_\phi t_d^3 + b_\phi t_d^2 + c_\phi t_d + d_\phi \tag{3.2a}$$

$$\overline{\theta}(t_d) = a_\theta t_d^3 + b_\theta t_d^2 + c_\theta t_d + d_\theta \tag{3.2b}$$

where $t_d = (t \mod 144)$ in units of 10 min, $a_\phi = 8.216 \times 10^{-5}$, $b_\phi = -2.316 \times 10^{-3}$, $c_\phi = -2.016 \times 10^{-2}$ and $d_\phi = 13.52$ for the ϕ times series and $a_\theta = -1.006 \times 10^{-5}$, $b_\theta = 5.616 \times 10^{-4}$, $c_\theta = -1.324 \times 10^{-2}$ and $d_\theta = 1.792$ for the θ times series.

If we observe the behaviour of $\overline{\phi}$ in Figure 3.2a, we can notice that the trading is heavier, i.e. we have a higher $\overline{\phi}$, at the beginning and end of the day than during the rest of the day. The opening and the closing of the NYSE are very peculiar times: they occur after and before the market is closed. This can explain the high volume-prices at the beginning and end of the day.

In the beginning of the day, the volume-price series has high values. This happens because our time series is not continuous. We have a time gap between the end of one day and the beginning of the next day. Information arrives during this period and the traders will have different opinions about what is going to happen in the morning. There is a lot of speculation. Therefore, when the Stock Market opens, there will be a huge quantity of money transactions. Notice that the volume-price is essentially the amount of money which is being traded. Everyone wants to sell or to buy fast so they can make more money than the others. Time is money so the faster you give your order, the higher profit you will have. As time passes by, the money traded will decrease. Now we do not have a urgency to sell or to buy.

However, after lunch time, we see an alteration in this pattern. Now the volume-price series starts to grow again. At the closing time, traders anticipate price changes overnight that can alter their portfolios, so they exchange larger amounts of money. Besides that, there are index mutual fund managers who need to make trades at the closing prices for administrative purposes and there are short sellers and hedgers who frequently close and hedge their positions at the end of the day. The volume-price increases at the end of the day at a higher rate than it decreases in the beginning of the day and the closing value is typically grater than the opening value. This may happen because there is a deadline (i.e. the closing time) that we do not have in the beginning of the day.

The volume traded has a bigger impact in the shape of the volume-price series than the price. Besides that, the series of volume-price inherit the oscillation-like structure from the series of trading volumes that we can see in Figure 1.1a. Rocha [1] showed that the correlation between the volume and the volume-price is approximately 0.8. Therefore we can admit that the pattern followed by ϕ and θ in the volume-price series is approximately the same that we would see in the volume series.

The U-Shaped pattern followed by ϕ is long known in the finances world [29, 30], it is very typical and it has been documented in various studies. This pattern is a characteristic of the volume series. However, we saw that the correlation between the volume and volume-price series is almost one. Therefore, it is expected that the volume-time series follows the same pattern. Various authors gave different reasons for the existence of this pattern. Admati and Pfleiderer [29] argue that high volume in a particular time segment reveals the presence of asymmetric information. Brock and Kleidon [31] defend that different trading strategies at the open and close of the markets form these volume patterns.

We can also see in Figure 3.2d the pattern followed by $\overline{\theta}$. We have to remember that θ accounts for the standard deviation of the volume-price series logarithm. In the beginning of the day there is a great variance in our data. This may happen because the traders have different perceptions about what is going to happen. As the time goes by, the standard deviation starts to decrease, first slowly and then faster. Finally, the standard deviation is very low at the end of the day. Apparently the traders exchange similar amounts of money. Maybe this happens because of the information received during the day. At the beginning of the day, people have different informations so they adopt different strategies. However, as the time passes, the information and the observation of the NYSE behaviour during that day leads the different traders to similar strategies.

These patterns, $\overline{\phi}$ and $\overline{\theta}$, happen every day and they are easily modelled by Equations (3.2). However, the fluctuations around these patterns, ϕ' and θ', need to be addressed more carefully

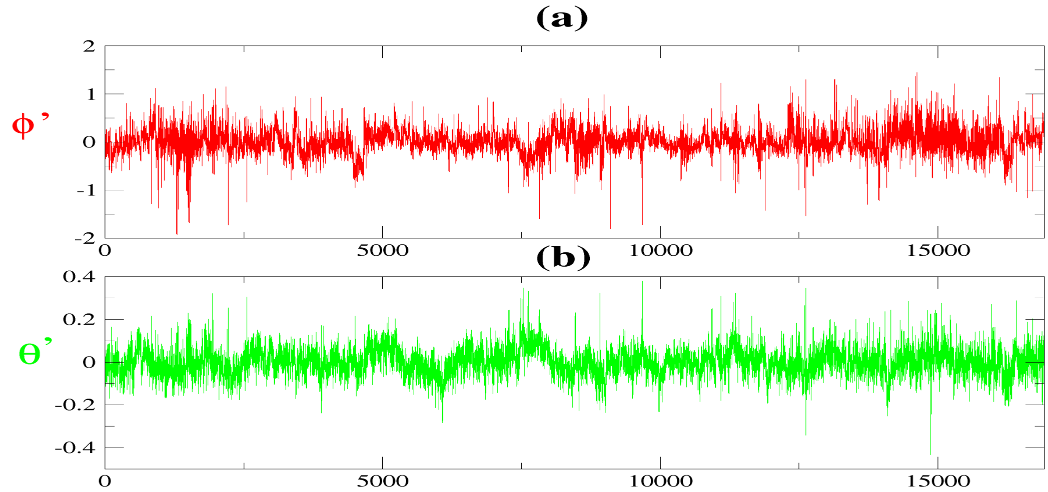

Figure 3.3: The fluctuations time series of the first 27 trading days: (a) ϕ' and (b) θ'.

since they have a strong stochastic behaviour. We will study the time series of the fluctuations in the next section.

3.2 Log-Normal Parameter Fluctuations

We got the time series of the fluctuations by subtracting the 20-day moving average pattern (Figure 3.2) from the data without the outliers (Figure 3.1). In each day, we subtracted the pattern from the 20 days before that day. We got Figure 3.3.

In order to make sure that we have removed all the periodicity from our time series, we build a power spectrum for both time series, before and after the process of removing the daily patterns, using the Fast Fourier Transform.

When the time series are viewed in the form of a frequency spectrum, certain aspects of the underlying processes are revealed. If the frequency spectrum include distinct periodic peaks, we may infer that the original processes have some kind of periodicity.

In Figures 3.4a and 3.4b we can see the power spectrum in a log-log scale. The solid lines are the power spectrum before we have removed the daily patterns. We can see some peaks that confirm our idea that the original time series are periodic. The power spectrum for the fluctuations is represented by the dots in the same picture. It is clear, for both time series, that the peaks observed in the solid lines disappeared. We have removed efficiently the periodicity from our time series.

We also computed the autocorrelation function, α, in a log-lin scale (see Figures 3.4c and 3.4d) for the ϕ' and θ' fluctuations time series to check the correlation between values of the series at different times. The ACF α has an exponential decay:

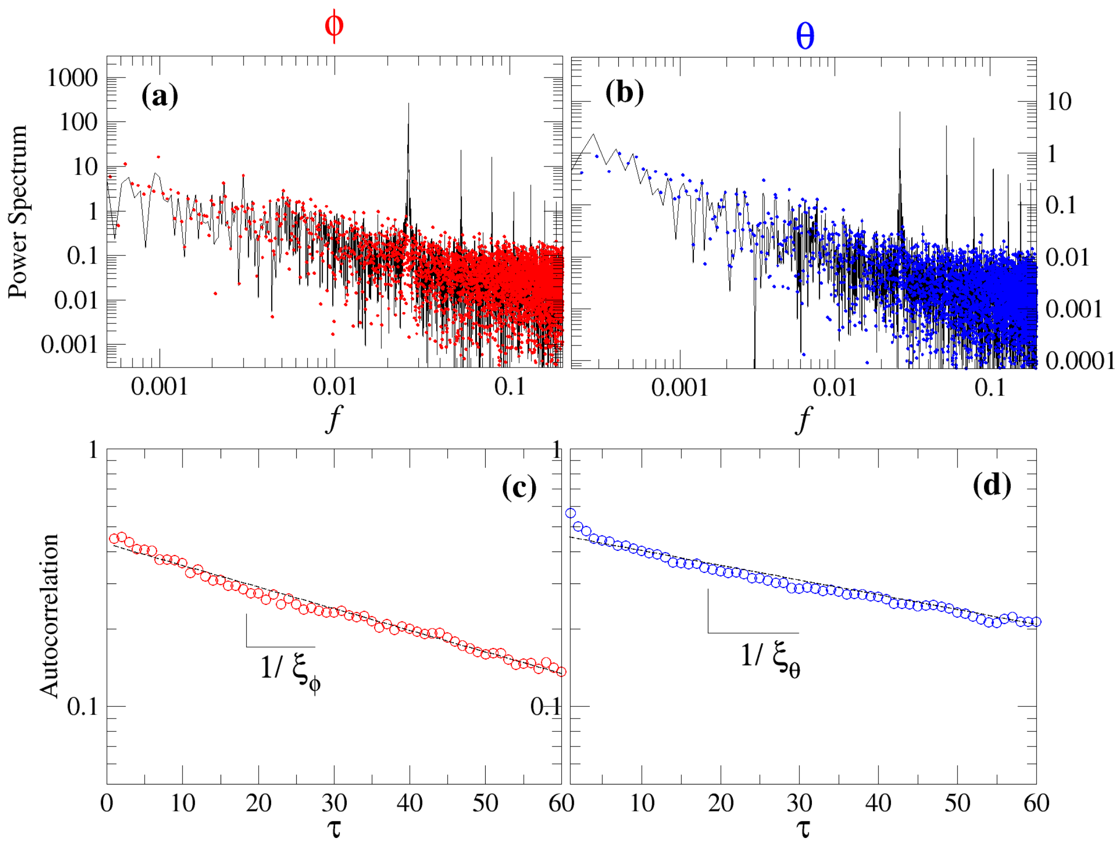

Figure 3.4: (a) The power spectrum of the ϕ and (b) θ time series is represented by the solid lines. The power spectrum of the (a) ϕ' and (b) θ' fluctuations is represented by the dots. (c) Autocorrelation function (dots) and linear function fitted to the data (dashed line) in a log-lin scale for the ϕ' time series and for the (d) θ' time series. $\xi_{\phi,\theta}$ represents the slope of the line fitted to the data.

$$\alpha_{\phi,\theta} = \beta_{\phi,\theta} e^{-\frac{\tau}{\xi_{\phi,\theta}}} \tag{3.3}$$

Therefore the logarithm of the ACF is a line. We fitted linear functions to both ACF. We got a R^2 bigger than 90% in both cases so we have a good fit. For ϕ', we have $\frac{1}{\xi_\phi} = 0.0192$ and $log(\beta_\phi) = -0.8496$. For θ', we have $\frac{1}{\xi_\theta} = 0.0132$ and $log(\beta_\theta) = -0.7776$. The $\frac{1}{\xi_{\phi,\theta}}$ constants 0.0192 and 0.0132 have units that are the inverse of the time because the exponential exponent cannot have units. Therefore, if we take the inverse of these constants, i.e. $\xi_\phi = 52.08$ and $\xi_\theta = 75.76$ we have the characteristic time for both series after which the process is uncorrelated, or, in other words, the characteristic time after which the processes have no memory.

We can see the probability density functions (PDF) of the fluctuations ϕ' and θ' in Figure 3.5. At first sight, these probability density functions appear to be normally distributed. We tested this hypothesis using the Shapiro-Wilk Normality test. According to Royston [32], an approximate p-value below 0.1 is enough to reject the null hypothesis of normality. For both time series, the Shapiro-Wilk test rejected the normality hypothesis.

Although the Shapiro-Wilk test rejectes the assumption of normality, we still plotted in Figure 3.5 the probability density function of the fluctuations ϕ' and θ' in a log-lin scale (dots) and the adjusted Gaussian PDF (dashed line), i.e., a Gaussian probability density function with the same mean and standard deviation as our data. We have a good fit for both time series in the

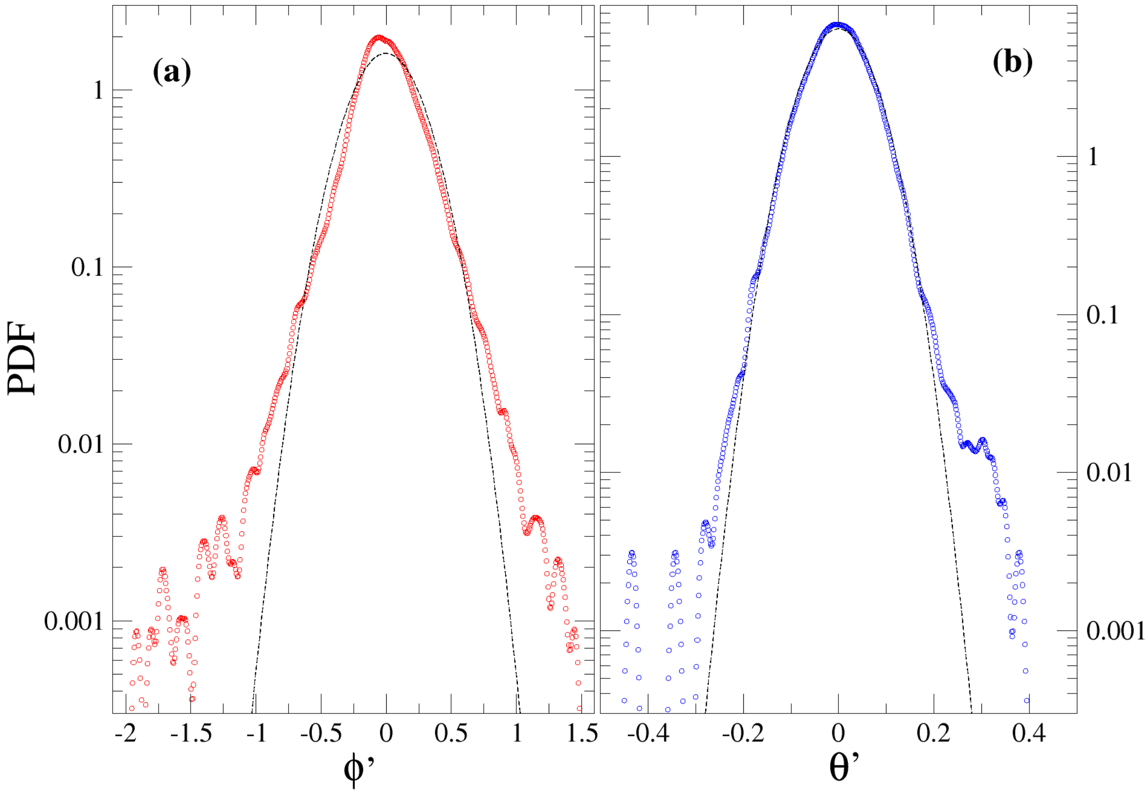

Figure 3.5: (a) Marginal probability density function (solid line) and Gaussian adjusted PDF (dashed line) of the fluctuations ϕ' and (b) θ' in a log-lin scale.

central region of the PDF. The marginal PDF of the θ' series has an almost perfect superposition with the Gaussian PDF. However, the fit in the tails is not so good. For the ϕ' time series, we also have a good fitting although it is not as good as for θ'. The skewness is almost zero in the empirical data. However, the kurtosis is approximately 6. This value is very high. Although the kurtosis has been associated to the "peakedness" of the distribution, Westafall [33] proved that the kurtosis has only an unambiguous interpretation in terms of tail extremity. A high kurtosis (like the one found in our data) draw our attention to the existence of outliers since there are outliers that contribute meaningfully to the computation of the kurtosis. In our data, we have fatter tails than it was expected. This can be related to the process of removing the outliers. It looks like some outliers have still remained in our data. Despite this, a Gaussian model is a good approach to study the evolution of the fluctuations ϕ' and θ'.

We have been studying each time series separately. However it is also important to see if there is any kind of correlation between our variables. We computed the covariance matrix Σ, which has components $\Sigma_{\phi\phi} = 0.0619$, $\Sigma_{\theta\theta} = 0.0039$, $\Sigma_{\phi\theta} = \Sigma_{\theta\phi} = -0.0036$. Besides this, we computed a correlation coefficient of -0.2311 between the two variables which show us that our variables are indeed correlated. Correlation and independence are different concepts. If we have zero correlation that does not imply independence. However, if the correlation is different from zero then we cannot have independence. In our case, the correlation coefficient is small. One can argue that it is almost zero. But one can also argue that the variables are negatively correlated.

In Figure 3.6a we plotted the joint probability density function of the fluctuations time series, ϕ' and θ'. In order to make a comparison, we also plotted in Figure 3.6b a multivariate normal

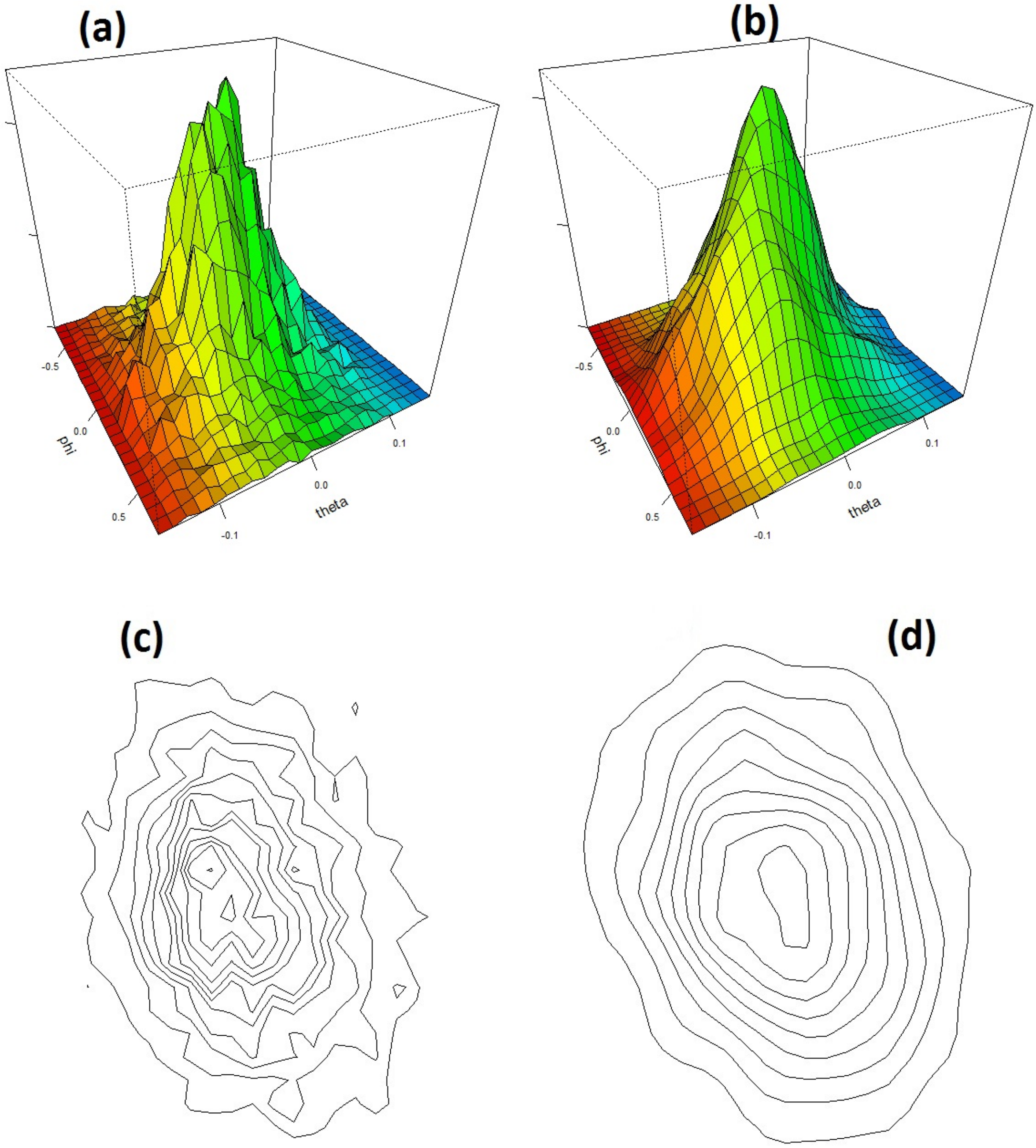

Figure 3.6: (a) Joint PDF of the empirical time series and (b) a multivariate normal distribution with mean vector and covariance matrix equal to the ones of our empirical data. (c) Contour plot for the ϕ' and θ' time series and (d) for a multivariate normal distribution with mean vector and covariance matrix equal to the ones of our empirical data.

distribution with mean vector and covariance matrix equal to the ones from the fluctuations time series. From these figures, we can see that the two plots are remarkably alike.

The PDF of the multivariate normal distribution represented in Figure 3.6b is:

$$p(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{3.4}$$

where $x = (\phi, \theta)$, $|\Sigma| = \Sigma_{\phi\phi}\Sigma_{\theta\theta} - \Sigma_{\phi\theta}^2$ is the determinant of the matrix Σ and μ is a 2-dimensional vector of zeros since the mean of ϕ' and θ' is approximately zero.

We also draw in Figures 3.6c the contour plots of the fluctuations and (d) the contour plots of the multivariate normal. We can notice that both contour plots area leaning towards the left which indicates a negative correlation between the variables. It is important to notice that the contour plot of the multivariate normal is not elliptical (as it was expected) because we simulated data from a multivariate normal and, after that, we build the joint distribution and the contour plot.

3.3 Markov Tests

To test the Markovianity of the series, we apply the Wilcoxon test to compare the distributions $p(x_1, \tau_1|x_2, \tau_2)$ and $p(x_1, \tau_1|x_2, \tau_2; x_3, \tau_3)$ for a fixed value of x_3. We are also going to infer the Markov length, τ_M.

In Figures 3.7 we can see that the values for the quotient t/t_0 are close to 1 for both time series for all values of τ. Thus, this suggests a Markov length $\tau_M = 600s = 10$min which is our time scale.

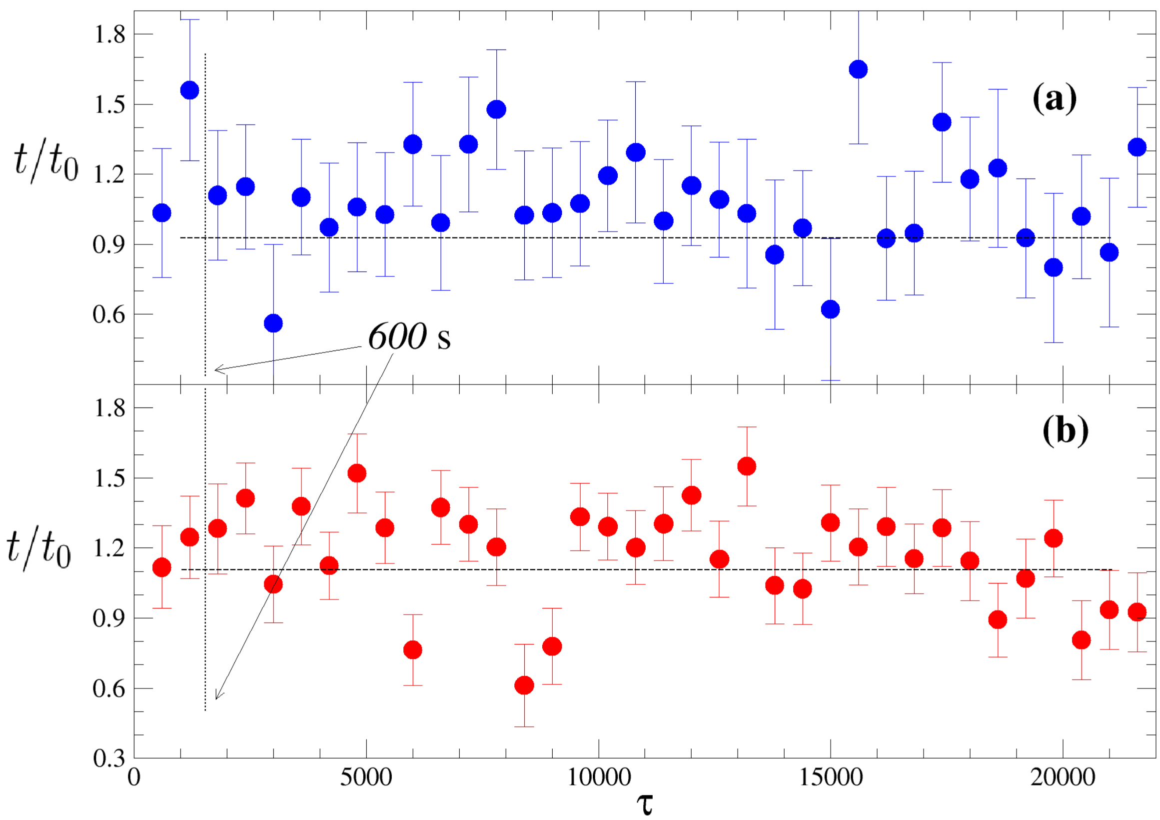

Figure 3.7: (a) Wilcoxon test to verify the Markovian property of the ϕ' time series and (b) the θ' time series, showing the Markov length of $\tau_M = 600s$.

Chapter 4

The Langevin Analysis

4.1 A Simple Model without Correlation

In this section we will implement the one dimensional Langevin Approach [34] to the two fluctuations time series represented in Figure 3.3. It is true that our time series are not independent. However, the correlation coefficient is small, namely $\rho = -0.2311$, so we will try to test a simple model in our data. Usually, the simplest models are the ones people use more often, even though they are not the ones that describe reality the best (one flagrant example is the Black-Scholes model).

We are going to assume this time series to be independent from each other. We use the routine “Langevin1D” applied to the fluctuations as described in Chapter 2. This routine retrieves $M^{(1)}$ and $M^{(2)}$ from Equations (2.59) and (2.61) which are essentially the means of the first and second conditional moment, in each bin and for each time step τ. We want to derive from our data the functions $D^{(1)}$ and $D^{(2)}$ that govern the following equations:

$$d\phi'(t) = D^{(1)}(\phi')dt + \sqrt{D^{(2)}(\phi')}dW_t \tag{4.1a}$$

$$d\theta'(t) = D^{(1)}(\theta')dt + \sqrt{D^{(2)}(\theta')}dW_t \tag{4.1b}$$

These are the derivatives with respect to time of $M^{(1)}$ and $M^{(2)}$. Thus, we computed the quotient $\frac{M^{(n)}}{\tau}$ [35] for each bin and for $n = 1, 2$. In Figure 4.1 we can see this quotient for the

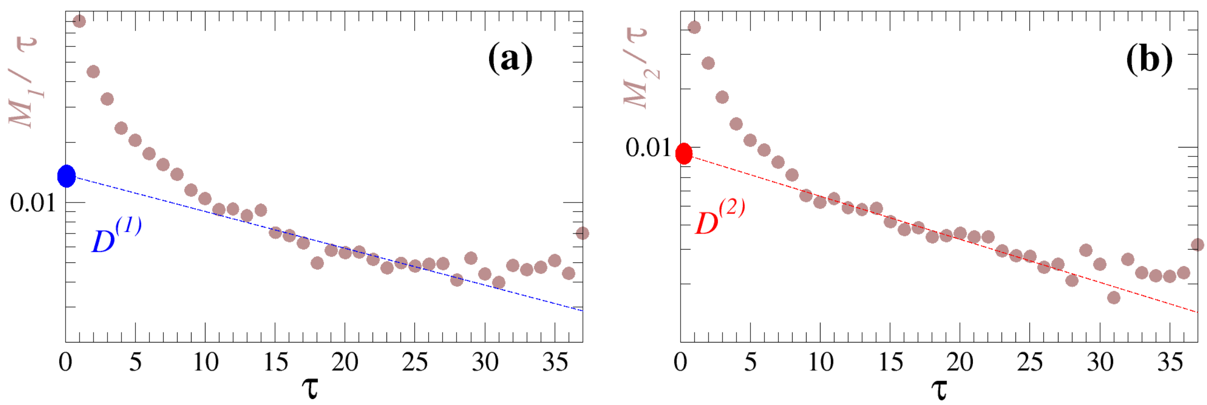

Figure 4.1: (a) Computation of $D^{(1)}$ and (b) $D^{(2)}$, in the bin that contains the mean, as the intersection of the linear fit with $\tau = 0$.

bin that contains the mean. We can see that $M^{(n)}$ is linear when we consider the time steps between 10 and 20. Thus, to find the derivatives with respect to time $D^{(n)}$ we just have to make a linear fit to $\frac{M^{(n)}}{\tau}$ (when the time steps are between 10 and 20) and see for which value that line intersects $\tau = 0$. That number is going to be our $D^{(n)}$ in that bin. If we do this for all the bins, we will have the $D^{(n)}$ function.

Note that we had to use this proxy because, apparently:

$$\lim_{\tau \to 0} \frac{M^{(n)}}{\tau} = +\infty \ . \tag{4.2}$$

This may happen because the time series have measurement noise which makes the empirical limit to diverge or simply because of round-off errors [36]. In order to surpass this, we use the fit to a linear region before the limit starts diverging.

Applying this methodology to all bins from our time series we get Figure 4.2. Now it is possible to write the Langevin equations for our time series using the values we obtained with the Langevin analysis as we can see in Equations (4.1), where $D_\phi^{(1)} = -0.08\phi$, $D_\phi^{(2)} = 0.006 + 0.07\phi^2$, $D_\theta^{(1)} = -8.0 \times 10^{-2}\theta$ and $D_\theta^{(2)} = 3.9 \times 10^{-4} - 7.9 \times 10^{-4}\theta + 5.6 \times 10^{-2}\theta^2$.

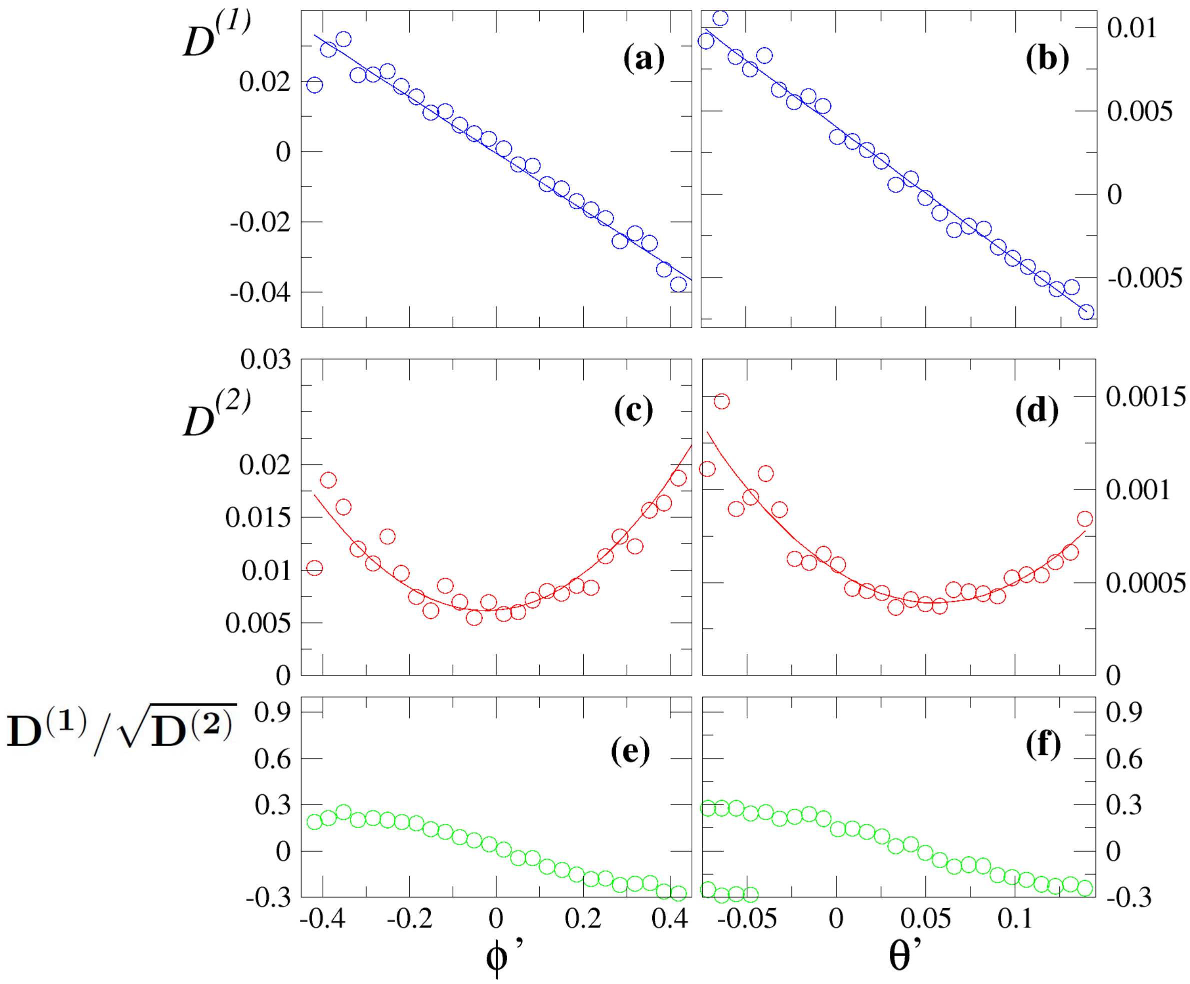

Figure 4.2: (a-d) D_1 and D_2 functions to the ϕ' (left) and θ' (right) time series. (e-f) Quotient $D^{(1)}/\sqrt{D_2}$ for ϕ' and θ', respectively.

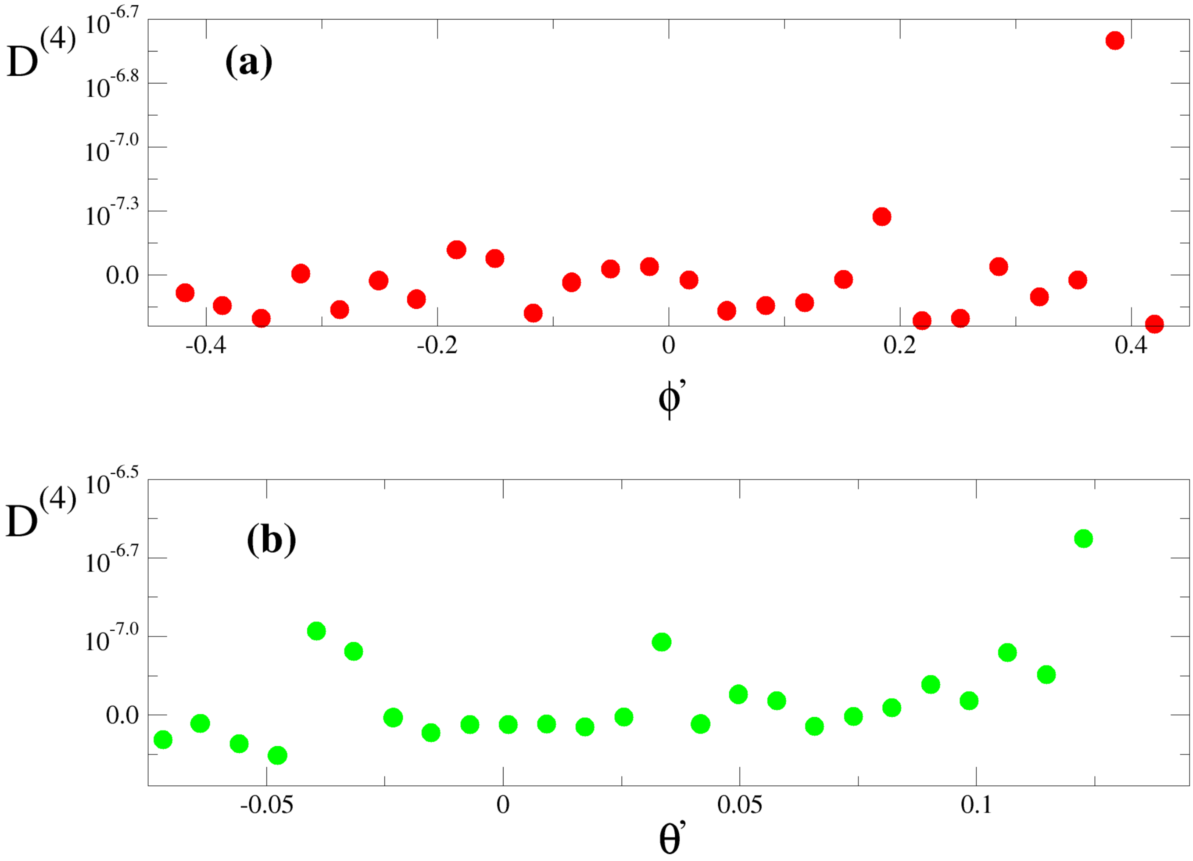

Figure 4.3: (a) D_4 function to the ϕ' and (b) θ' time series.

The $D^{(1)}$ coefficient is linear for both the ϕ' and θ' time series. This means we have an oscillator with a string constant that corresponds to the line's slope. The string has a fixed point in zero since the line contains the origin of the referential.

In Figure 4.2e and Figure 4.2f we plotted the quotient $D^{(1)}/\sqrt{D^{(2)}}$ to check if the order of magnitude of the functions that govern the deterministic and stochastic part of Equations (4.1) is the same. That quotient varies between -0.3 and 0.3 for both time series, which means that $\sqrt{D^{(2)}}$, i.e., the coefficient of the stochastic part, has a heavier weight in the Equations (4.1) than $D^{(1)}$, the coefficient of the deterministic part.

We also computed the $D^{(4)}$ function for both time series. The results obtained are in Figure 4.3. Here we can see that $D^{(4)}$ is almost zero for both time series. According to the Pawula Theorem [37], if $D^{(4)}$ is zero or very small when compared to $D^{(1)}$ and $D^{(2)}$, then we can stop the Kramers-Moyal expansion at $n = 2$ since the terms of bigger order are zero.

4.2 Modelling the Coupling between ϕ' and θ'

Since our time series are not independent, we are going to perform the Langevin Analysis 2D described in Chapter 2, i.e., assuming there exists dependence between the two time series.

The routine gives us the values of $M^{(1)}$ and $M^{(2)}$, like we saw in the previous section. Thus, we need to compute $D^{(1)}$ and $D^{(2)}$. However, we do not have a linear behaviour of $\frac{M^{(n)}}{\tau}$ like we had when we assumed that ϕ' and θ' were independent from each other. This may happen because our sampling frequency is too high. We are going to use as a proxy the value of $M^{(n)}(x)$ in $\tau = 1$.

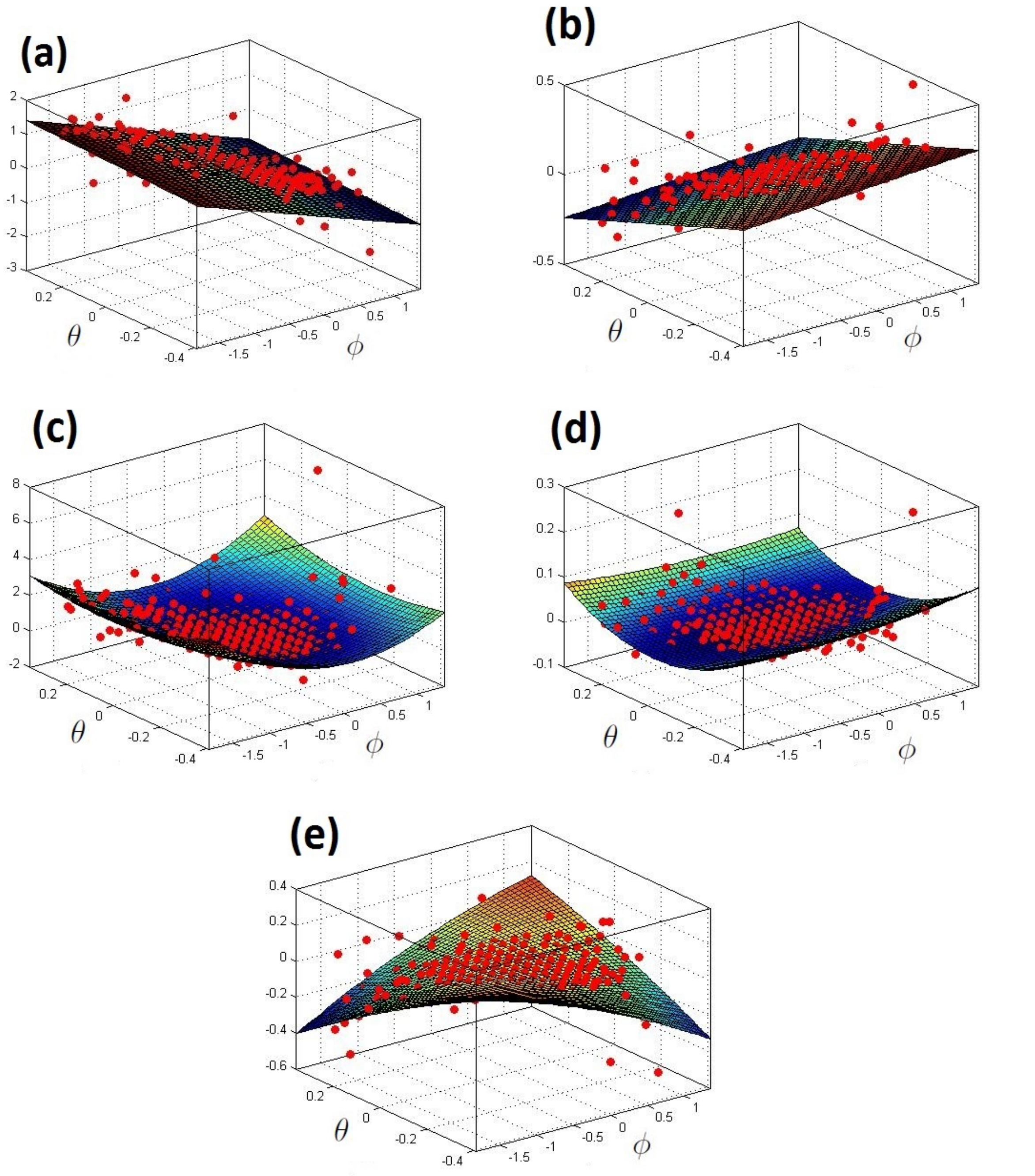

Figure 4.4: Functions from the drift vector and diffusion matrix obtained by the routine Langevin2D: (a) $D^{(1)}_{\phi}$, (b) $D^{(1)}_{\theta}$, (c) $D^{(2)}_{\phi\phi}$, (d) $D^{(2)}_{\theta\theta}$, (e) $D^{(2)}_{\phi\theta}$.

An interval of 10 minutes in the NYSE is huge, considering there are transactions happening in a microsecond scale. When we have a Markov process, the conditional moments change in a linear way to small values of τ. Then, for larger values of τ, they start deviating from linearity.

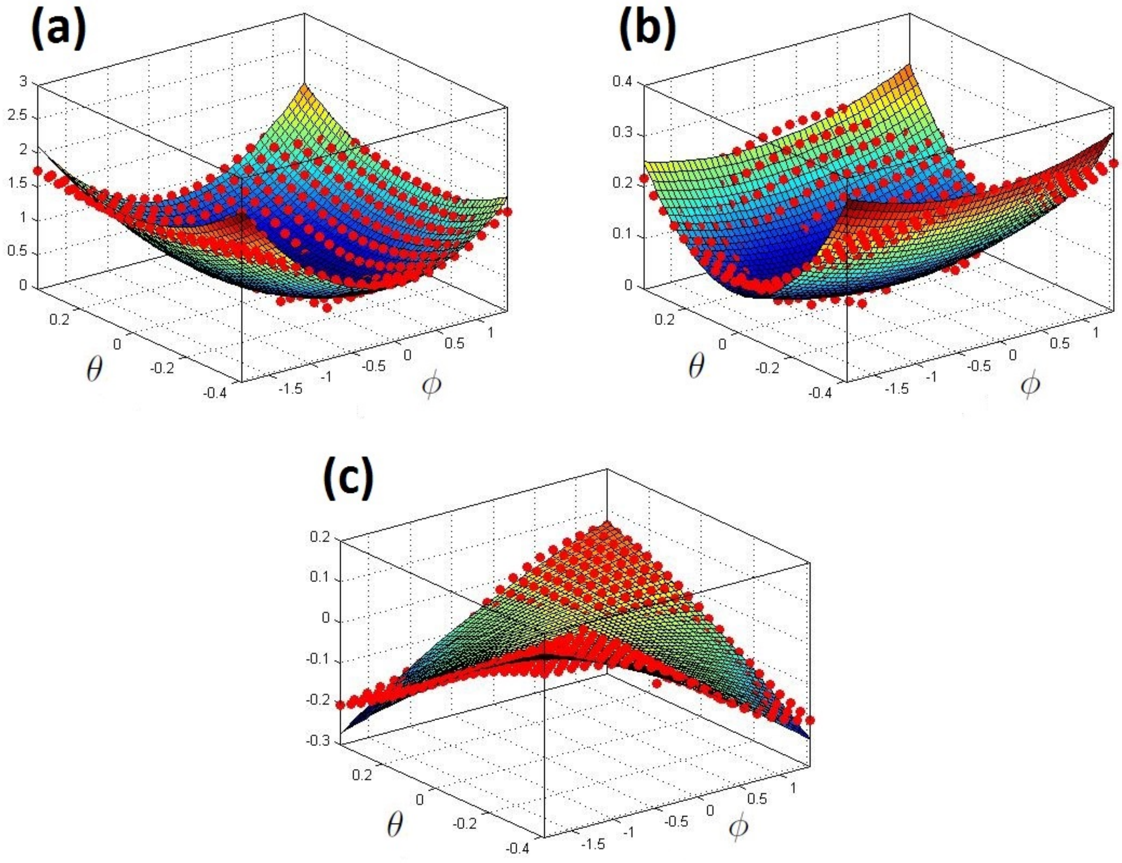

Figure 4.5: The three components of the g matrix: (a) $g_{\phi\phi}$, (b) $g_{\theta\theta}$, (c) $g_{\phi\theta}$.

This may be happening in our data but we cannot notice because of the frequency of our time series. Maybe we have this linear behaviour for a $\tau < 10$min. But our $\tau_1 = 10$min so we do not notice this linear behaviour. One way of getting over this problem is to approximate the $D^{(n)}(x)$ by the value of $M^{(n)}(x)$ in $\tau = 1$.

We can be making an approximation error (or not). We do not know how the variation takes place in the linear part. If the linear part is immediately below the 10 minutes, then we should have a good approximation. However, if the linear variation is in the order of some seconds, then the first 10 minutes are very far from the linear part and our approximation may not be very good. Nevertheless, this is the best estimative that we can make with this data. It is better than to use the other values of τ.

In Figure 4.4 we plotted the functions from the drift vector and diffusion matrix (notice that this matrix is symmetric) obtained by the routine Langevin2D [34]. For simplicity we are going to suppress the prime symbol noticing that we only address the fluctuations. Thus, in the two dimensional case we have a system of coupled Langevin equations that govern the evolution of ϕ and θ:

$$\begin{bmatrix} d\phi(t) \\ \\ d\theta(t) \end{bmatrix} = \begin{bmatrix} {D_\phi}^{(1)}(\phi,\theta) \\ \\ {D_\theta}^{(1)}(\phi,\theta) \end{bmatrix} dt + \begin{bmatrix} g_{\phi\phi}(\phi,\theta) & g_{\phi\theta}(\phi,\theta) \\ \\ g_{\phi\theta}(\phi,\theta) & g_{\theta\theta}(\phi,\theta) \end{bmatrix} \begin{bmatrix} dW_t^{(1)} \\ \\ dW_t^{(2)} \end{bmatrix} \tag{4.3}$$

where $gg^T = D^{(2)}$ with:

$$D^{(2)} = \begin{bmatrix} D^{(2)}_{\phi\phi} & D^{(2)}_{\phi\theta} \\ \\ D^{(2)}_{\phi\theta} & D^{(2)}_{\theta\theta} \end{bmatrix}. \tag{4.4}$$

We fitted polynomials of degree one to the functions from the drift vector and polynomials of degree two to the functions from the diffusion matrix:

$$D^{(1)} \approx a + b\phi + c\theta \tag{4.5a}$$

$$D^{(2)} \approx a + b\phi + c\theta + d\phi^2 + e\phi\theta + f\theta^2 \tag{4.5b}$$

whose coefficients are presented in the next table:

Term	1	ϕ	θ	ϕ^2	$\phi\theta$	θ^2
$D^{(1)}_{\phi}$	-0.0085	-0.7143	0.2812			
$D^{(1)}_{\theta}$	-0.0031	0.0293	-0.5023			
$D^{(2)}_{\phi\phi}$	-0.1233	0.1107	0.3563	0.9013	0.7350	5.8615
$D^{(2)}_{\theta\theta}$	-0.0017	0.0037	-0.0104	0.0059	-0.0186	0.5253
$D^{(2)}_{\phi\theta}$	-0.0081	-0.0046	-0.0267	-0.0222	0.4385	-0.1901

Table 4.1: Coefficients for the $D^{(1)}$ and $D^{(2)}$ functions, obtained by the Langevin analysis.

In order to have all of the coefficients of Equation (4.3), we have to compute the g matrix. This matrix g is not unique. If we have an orthogonal matrix Q and $gg^T = D^{(2)}$ then $h = gQ$ also satisfies $hh^T = D^{(2)}$ since

$$hh^T = (gQ)(gQ)^T = gQQ^Tg^T = gg^T = D^{(2)} \tag{4.6}$$

For now, we will just compute one possible matrix that satisfies this property. If the matrix $D^{(2)}$ is diagonalizable then we know that there exists an invertible matrix P satisfying:

$$PD^{(2)}P^{-1} = d \tag{4.7}$$

where d is a diagonal matrix that contains the eigenvalues of $D^{(2)}$ in the main diagonal and P is a matrix whose columns are the eigenvectors of $D^{(2)}$. Now it is easy to show that

$$g = P^{-1}\sqrt{d}P \tag{4.8}$$

where $\sqrt{d}$ is the matrix obtained after taking the square root of each element of the diagonal matrix d. Notice that:

$$gg^T = (P^{-1}\sqrt{d}\,P)(P^{-1}\sqrt{d}\,P)^T = P^{-1}\sqrt{d}\,PP^{-1}\sqrt{d}\,P = P^{-1}dP = D^{(2)} \tag{4.9}$$

since $P^{-1} = P^T$ because symmetric matrices have orthogonal eigenvectors.

After implementing this procedure we obtained the three components of the g matrix (notice that this matrix is symmetric) and we fitted quadratic forms to this components:

$$g \approx a + b\phi + c\theta + d\phi^2 + e\phi\theta + f\theta^2 \tag{4.10}$$

whose coefficients are presented in the next table:

Term	1	ϕ	θ	ϕ^2	$\phi\theta$	θ^2
$g_{\phi\phi}$	0.2185	0.0918	0.2255	0.4850	0.2925	4.0541
$g_{\theta\theta}$	0.0360	0.0174	-0.0128	0.0210	0.0245	1.5197
$g_{\phi\theta}$	-0.0111	-0.0051	-0.0158	-0.0134	0.2936	-0.1835

Table 4.2: Coefficients for the g functions.

We can see in Figure 4.5 the empirical results obtained and the surfaces fitted to this data.

Chapter 5

Approaching Non-Stationarity

After introducing our framework in Chapters 3 and 4 and applying it to the volume-price series, we now deduce the formula of all the moments $\mathbb{E}[s^n], n = \{0, 1, 2, ...\}$ of the log-normal distribution. We are going to use the notation $\langle s^n \rangle$ with the same meaning as $\mathbb{E}[s^n]$. It is a well known statistical result [38] that, if we have all the moments from a distribution, we can deduce its probability density function using a Fourier transform. It is possible to have a closed-form for the n^{th}-moment of s, since s follows a log-normal distribution whose PDF is given by Equation (3.1). The moments $\langle s^n \rangle, n = \{0, 1, 2, ...\}$ of our distribution are given by:

$$\langle s^n \rangle = \int_{-\infty}^{+\infty} s^n p(s, \phi(t), \theta(t)) ds = e^{n\phi + \frac{n^2\theta^2}{2}} \ . \tag{5.1}$$

Assuming that all time dependency is incorporated in the distribution parameters ϕ and θ one is able to fully characterize the non-stationary time series of the volume-price: one just needs to model the ϕ and θ evolution through time. Indeed, these parameters are the sum of a daily average pattern $\overline{\phi}$ and $\overline{\theta}$ with the fluctuations ϕ' and θ':

$$\phi = \overline{\phi} + \phi' \tag{5.2a}$$

$$\theta = \overline{\theta} + \theta' \tag{5.2b}$$

The daily patterns $\overline{\phi}$ and $\overline{\theta}$ are the ones observed in Figure 3.2 and can be described by the expressions in Equations (3.2). Since we already have the expressions for the daily patterns $\overline{\phi}$ and $\overline{\theta}$, now we need to describe the fluctuations ϕ' and θ'. In our model we are going to assume that these fluctuations obey the system of Langevin equations given in Equation (4.3). For simplicity we are going to suppress the prime symbol noticing that we only address the fluctuations.

In order to integrate these equations for the parameters fluctuations, we will have to use their discrete versions in the Itô's description, namely:

$$\phi(t + \Delta t) = \phi(t) + D_\phi^{(1)}(\phi(t), \theta(t))\Delta t + g_{11}(\phi(t), \theta(t))\sqrt{\Delta t}\ r_1 + g_{12}(\phi(t), \theta(t)\sqrt{\Delta t}\ r_2 \tag{5.3a}$$

$$\theta(t + \Delta t) = \theta(t) + D_\theta^{(1)}(\phi(t), \theta(t))\Delta t + g_{21}(\phi(t), \theta(t))\sqrt{\Delta t}\ r_1 + g_{22}(\phi(t), \theta(t))\sqrt{\Delta t}\ r_2 \tag{5.3b}$$

where r_1 and r_2 are random numbers from a Gaussian distribution with mean zero and standard deviation 1.

Having generated a sample of fluctuations, we then add the daily patterns in Equations (3.2)

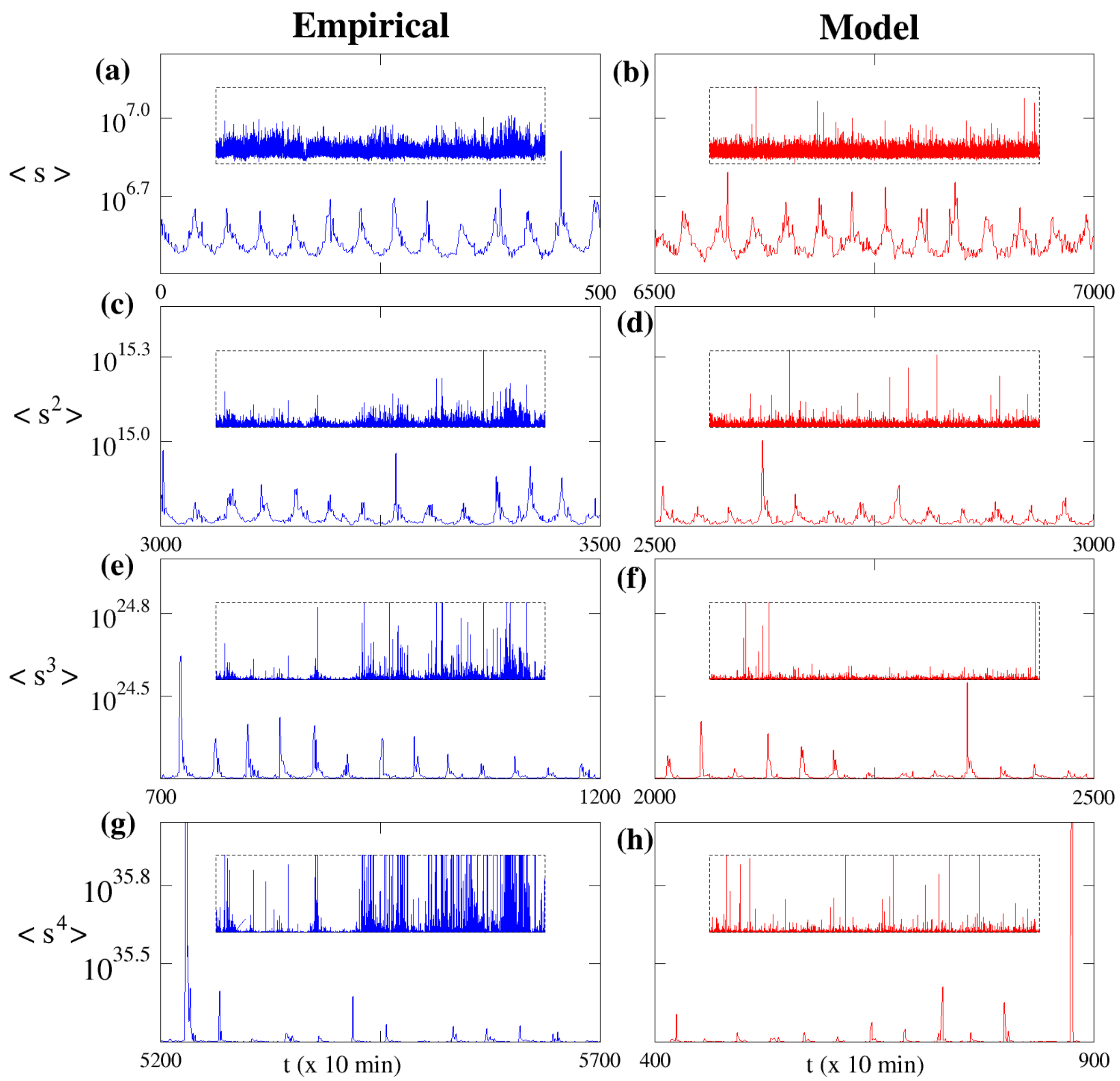

Figure 5.1: (a) Time series of empirical $\langle s \rangle$ and (b) corresponding modelled series. Therefore, (c-d) $\langle s^2 \rangle$, (e-f) $\langle s^3 \rangle$ and (g-h) $\langle s^4 \rangle$. Inside each plot, there is a sub-plot with the corresponding entire series.

and obtain the modelled time series for ϕ_{mod} and θ_{mod}. Inserting ϕ_{mod} and θ_{mod} in Equation (5.1) yields the modelled volume-price moments.

We next compare the modelled time series of the first four moments $\langle s^n \rangle$, $n = 1, ..., 4$, with the empirical ones which are obtained by replacing in Equation (5.1) the original time series of ϕ and θ represented in Figure 3.1.

In Figure 5.1 we have the series obtained from our model versus the empirical ones, for the first four moments. If we look to the entire time series, we can see that for $n = 1, 2$, our model can explain the extreme events. However, for $n = 3, 4$, we have far more extreme events in the empirical series than in the modeled ones. When we look to the zoom in Figure 5.1, we can see the same pattern in both series. In our model, the extreme events do not happen at the same time as in the empirical ones, but they will eventually happen.

In Figure 5.2 we have the empirical and modelled probability density functions for the fluctuations of ϕ and θ. We can see that our model can explain better the ϕ fluctuations than the θ fluctuations. In Figure 5.2a we see that the densities are quite close, specially in the central

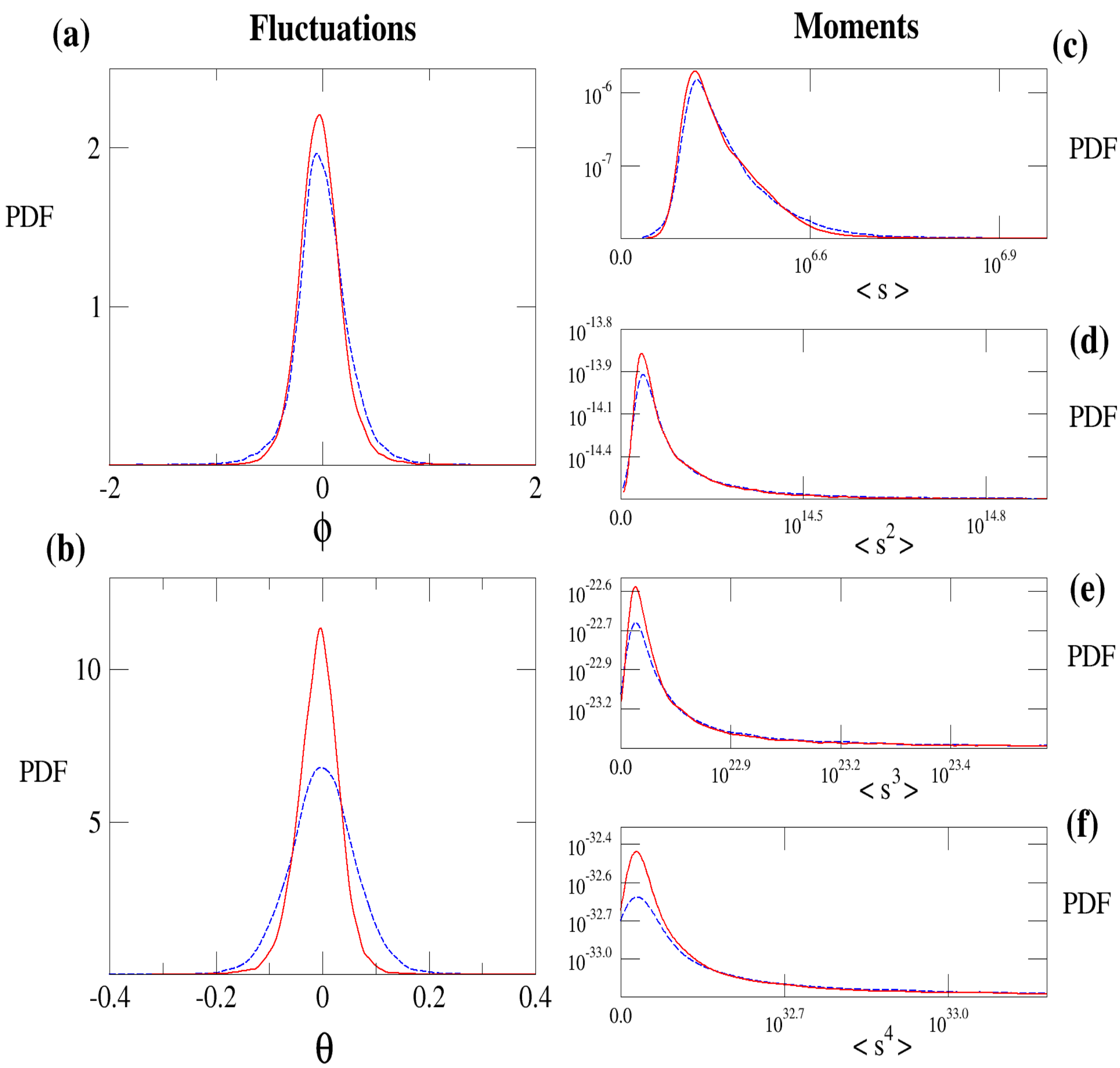

Figure 5.2: (a) PDFs for the empirical fluctuations of ϕ (dashed line) and for the modelled fluctuations (solid line). (b) PDFs for the empirical fluctuations of θ (dashed line) and for the modelled fluctuations (solid line). (c-f) PDFs for the $\langle s^n \rangle$ time series. The dashed line represents the empirical PDF and the solid line is the PDF obtained from our model.

region. However, in Figure 5.2b we do not have a fit as good since it is easier to model means than standard deviations. This can be explained noticing that ϕ is a first-order moment while θ is the square root of a second-order central moment.

In Figure 5.2(c-f) we plotted the empirical and theoretical probability distributions of $\langle s^n \rangle, n = 1, ..., 4$. Our model has a good fit in the first moments and can be used to model them since the theoretical and empirical distributions are very close to each other. Since the ϕ parameter is better modelled than the θ parameter, it is expected that for higher moments, when θ is dominant over ϕ, we do not achieve such good results.

Chapter 6

Discussion and Conclusions

The main goal of this dissertation was to model the non-stationary time series of the volume-price. By assuming that the log-normal had the best fit to the data in each 10-minutes window, this goal resumes to the one of studying the parameters ϕ and θ of this distribution, which are themselves stochastic variables. We were able to show that we can describe the time series of these parameters by decomposing the variables as a sum of two terms: one accounting for the daily pattern and another regarding the fluctuation around that average pattern. The fluctuations are modelled using a system of Langevin equations whose coefficients we retrieved from our empirical data. From here, we proposed a framework to reconstruct the evolution of all the moments of the volume-price distribution.

Our model reproduces well the first moments, being therefore suitable to study the volume-price of the stocks from the NYSE. It would be interesting if, in the future, we conduct tests to see if this framework can be used to make previsions about the evolution of the volume-price. In particular, it could be a suitable approach for the calculation of the *Value at Risk* (VaR) of a stock's portfolio.

This work leaves some open questions to be answered. It is true that we achieved a good model to the ϕ fluctuations, but we could not match this result to the θ fluctuations. One possible explanation is related with the outliers: we removed all the points which did not lie in a 5σ interval from the mean. However, when we plotted the time series without the outliers, there were still some extreme values that look more like measurement errors than fluctuations as we can see in Figure 3.1. We chose to use the 5σ criterion because we tried to minimize the number of points taken from our sample in order to let our data as close as possible to the original one. However, if one prefers to choose a stricter criteria, like using a 3σ interval, then the time series would have lesser outliers and maybe the results would be more easily modelled.

There are many models in the literature that enable us to study and to model stochastic time series such as autoregressive models [39], moving-average models [40] and autoregressive integrated moving average models [41]. One may ask, why did we choose the Langevin model instead of all the others. One strong argument in favour of this model is that it not only allows us to describe the evolution of our time series, but it may also give us an equation, Fokker-Planck equation, to describe the evolution of the volume-price distribution. Further work should be done in trying to extract such an equation from the equations we already have. If one is able to do this, then we would have much more information about the volume-price evolution and we could apply this information to the computation of the *Value at Risk* or other risk measures.

A comparison between our model and the classic models that have been used for studying time series would be an interesting work to develop in the future. It is true that our model has the advantage already stated of being able to produce an equation to the evolution of the distribution of the volume price. However, the results achieved by our model may be indeed better than the ones achieved by the classical models. In order to test this hypothesis, we should do this comparative study.

Another question that raises from the main findings of this dissertation is related with the sampling frequency of our data. In the NYSE, where transactions are happening at the microsecond scale, 10 minutes is a very large interval. If we were able to redo this analysis with a smaller sampling frequency, maybe we would be able to extract stochastic differential equations which explain better the behaviour of the fluctuations in our data.

We proposed a model that is capable of modeling the central region of the volume-price distribution. However, it does not have a good fit in the tails. It would be interesting to develop a two dimensional model to describe the behaviour of the tails. Rocha [1] had already described the tails of the volume-price as a one parametric inverse gamma distribution. But if one is able to fit a two parametric model to the tails, then it may yield better results. After this, it would be a good idea to combine these two models and see if we get a better fit to the real data.

Finally, this work gave us important insight in the study of non-stationary time series and we have proposed here a methodology that is going to be very useful in numerous fields. This framework is general enough to be applied to other markets besides the NYSE and also to other fields of study like physiology, when we are trying to study the heart interbeat intervals or geology, in order to study seismic time series.

Bibliography

[1] P. Rocha, "Stochastic evolution of parameters defining probability density functions: Application to the new york stock market," Master's thesis, Faculty of Sciences of Lisbon University, 2014.

[2] H. Furstenberg, *Stationary Processes and Prediction Theory.* Princeton University Press, 1960.

[3] P. Rocha, F. Raischel, J. Cruz, and P. Lind, in *3rd SMTDA Conference Proceedings*, 2015, pp. 619–627.

[4] P. Rocha, F. Raischel, J. Boto, and P. Lind, *Journal of Physics: Conference Series*, vol. 574, 012148, 2014.

[5] R. Friedrich, J. Peinke, M. Sahimi, and M. Tabar, "Approaching complexity by stochastic methods: From biological systems to turbulence," *Physical Review*, vol. 506, p. 87, 2011.

[6] P. Laplace, *A philosophical essay on probabilities.* New York: J. Wiley, 1902.

[7] R. Brown, "A brief account of microscopical observations made in the months of june, july and august, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies." *Philosophical Magazine*, vol. 4, p. 161–173, 1828.

[8] L. Bachelier, "Théorie de la spéculation," *Annales Scientifiques de l' École Normale Supérieure*, vol. 3, no. 17, pp. 21–86, 1900.

[9] A. Einstein, "Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen," *Annalen der Physik*, vol. 17 (8), p. 549–560, 1905.

[10] D. Lamberton and B. Lapeyre, *Introduction to Stochastic Calculus Applied to Finance.* Chapman and Hall/CRC, 2008.

[11] F. Beichelt, *Stochastic Processes in Science, Engineering and Finance.* New York: Chapman and Hall/CRC, 2006.

[12] M. Vetterli, J. Kovacevic, and V. Goyal, *Foundations of Signal Processing.* Cambridge University Press, 2014.

[13] J. Hunter, *Lecture Notes on Applied Mathematics: Methods and Models.* University of California, 2009. [Online]. Available: www.math.ucdavis.edu/~hunter/m280_09/ch.pdf

[14] F. Yilmaz, H. Öz, and G. Weber, “Itô-Taylor expansions for systems of stochastic differential equations with numerical applications,” 2013.

[15] L. Arnold, *Stochastic Differential Equations: Theory and Applications.* New York: J. Wiley, 1974.

[16] F. Black and M. Scholes, “The pricing of options and corporate liabilities,” *Journal of Political Economy*, vol. 81(3), pp. 637–654, 1973.

[17] R. Merton, “Theory of rational option pricing,” *The Bell Journal of Economics and Management Science*, vol. 4(1), pp. 141–183, 1973.

[18] J. Jackwerth and M. Rubinstein, “Recovering stochastic processes from option prices,” *Contemporary Studies in Economic and Financial Analysis*, vol. 94, 2012.

[19] R. Schmalensee and R. Trippi, “Common stock volatility expectations implied by option premia,” *Journal of Finance*, vol. 33, pp. 129–147, 1978.

[20] P. Dennis and S. Mayhew, “Risk-neutral skewness: Evidence from stock options,” *Journal of Financial and Quantitative Analysis*, vol. 37, pp. 471–493, 2002.

[21] J. Cox, “Notes on option pricing I: Constant elasticity of variance diffusions,” *Unpublished Draft*, 1975.

[22] S. Heston, “A closed-form solution for options with stochastic volatility with applications to bond and currency options,” *The Review of Financial Studies*, vol. 6(2), pp. 327–343, 1993.

[23] C. Beck, “Superstatistics: Theoretical concepts and physical applications,” in *Anomalous Transport.* Wiley-VCH Verlag GmbH and Co. KGaA, 2008.

[24] R. Friedrich and J. Peinke, “Description of a turbulent cascade by a fokker-planck equation,” *Physical Review Letters*, vol. 78 (863), 1997.

[25] R. Stresing, D. Kleinhans, R. Friedrich, and J. Peinke1, “Different methods to estimate the Einstein-Markov coherence length in turbulence,” *Physical Review E*, vol. 83, 046319, 2011.

[26] C. Renner, J. Peinke, and R. Friedrich, “Experimental indications for markov properties of small-scale turbulence,” *Journal of Fluid Mechanics*, vol. 443, pp. 383–409, 2001.

[27] S. Shapiro and M. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965. [Online]. Available: http://biomet.oxfordjournals.org/content/52/3-4/591.short

[28] P. Rocha, F. Raischel, J. Boto, and P. Lind, “Uncovering the evolution of nonstationary stochastic variables: The example of asset volume-price fluctuations,” *Physical Review E*, vol. 93,052122, May. [Online]. Available: link.aps.org/doi/10.1103/PhysRevE.93.052122

[29] A. Admati and P. Pfleiderer, “A theory of intraday patterns: Volume and price variability,” *The Review of Financial Studies*, vol. 1, pp. 3–40, 1988.

[30] P. Jain and G. John, "The dependence between hourly prices and trading volume," *Journal of Financial and Quantitative Analysis*, vol. 23, pp. 269–284, 1986.

[31] W. Brocka and A. Kleidon, "Periodic market closure and trading volume: A model of intraday bid and asks," *Journal of Economic Dynamics and Control*, vol. 16, pp. 451–489, 1991.

[32] J. Royston, "Remark AS R94: A remark on algorithm AS 181: The W test for normality," *Applied Statistics*, vol. 44, pp. 547–551, 1995.

[33] P. Westfall, "Kurtosis as peakedness, 1905-2014 R.I.P." *The American Statistician*, vol. 68, pp. 191–195, 2014.

[34] P. Rinn, P. Lind, M. Wachter, and J. Peinke, "Langevin: An R package for stochastic data analysis," *Journal of Open Research Software*, 2016.

[35] P. Lind, M. Haase, F. Böttcher, J. Peinke, D. Kleinhans, and R. Friedrich, "Extracting strong measurement noise from stochastic series: applications to empirical data," *Physical Review E*, vol. 81, 041125, 2010.

[36] E. Darulová, "Programming with numerical uncertaintie," Ph.D. dissertation, École Polytechinque Fédérale de Lausanne, 2014.

[37] H. Risken, *Fokker-Planck Equation.* Berlin: Springer, 1984.

[38] S. Rabbani, "Probability density function in terms of moments," 2008. [Online]. Available: http://srabbani.com/moments2.pdf

[39] G. Yule, "On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London*, vol. 226, p. 267–298, 1927.

[40] W. Enders, *Stationary Time-Series Models. Applied Econometric Time Series.* New York: J. Wiley, 2004.

[41] J. Contreras, R. Espinola, F. Nogales, and A. Conejo, "Arima models to predict next-day electricity prices," *Institute of Electrical and Electronics Engineers(IEEE) Transactions on Power Systems*, vol. 18, pp. 1014–1020, 2003.